Taimoor Ali (2k20/ITE/136)

Fayaz Ali (2k20/ITE/39)

Mubeen Gul (2k20/ITE/152)

Rahul Dev (2k20/ITE/102)

Ali Kolachi (2k20/ITE/17)

Assignment : Data Science

BSIT-PART 3RD SEMISTER-6TH

To improve data quality, start by assessing your data and defining what is

acceptable quality. Organizations also need to eliminate data silos, correct

errors upfront, and promote a data-driven culture. The best way to improve

data quality is to use an automated data quality management solution, such

as DataBuck.

How can we assess our data quality :-

Data quality meets six dimensions: accuracy, completeness, consistency,

timeliness, validity, and uniqueness.

How to Improve Data Quality in our Organization

1-Assess Your Data.

2-Define Acceptable Data Quality.

3-Correct Data Errors Up Front.

4-Eliminate Data Silos.

5-Make Data Accessible to All Users.

6-Use the Correct Data.

7-Impose a Defined Set of Values for Common Data.

8-Secure Your Data.

How can we improve the quality of data in an experiment?

You can increase the validity of an experiment by controlling more

variables, improving measurement technique, increasing randomization to

reduce sample bias, blinding the experiment, and adding control or placebo

groups.

What are the 4 categories of data quality?

There are five traits that you'll find within data quality: accuracy,

completeness, reliability, relevance, and timeliness.

Github link (Code)

## Data Quality

1. [Reliable and Scalable Data Ingestion at Airbnb](https://www.slideshare.net/HadoopSummit/reliable-and-scalable-data-ingestion-at-airbnb-63920989) `Airbnb` `2016`

2. [Monitoring Data Quality at Scale with Statistical Modeling](https://eng.uber.com/monitoring-data-quality-at-scale/) `Uber` `2017`

3. [Data Management Challenges in Production Machine Learning](https://research.google/pubs/pub46178/) ([Paper](https://thodrek.github.io/CS839\_spring18/papers/p1723-polyzotis.pdf)) `Google` `2017`

4. [Automating Large-Scale Data Quality Verification](https://www.amazon.science/publications/automating-large-scale-data-quality-verification) ([Paper](https://assets.amazon.science/a6/88/ad858ee240c38c6e9dce128250c0/automating-large-scale-data-quality-verification.pdf))`Amazon` `2018`

5. [Meet Hodor — Gojek’s Upstream Data Quality Tool](https://www.gojek.io/blog/meet-hodor-gojeks-upstream-data-quality-tool) `Gojek` `2019`

6. [An Approach to Data Quality for Netflix Personalization Systems](https://www.youtube.com/watch?v=t7vHpA39TXM) `Netflix` `2020`

7. [Improving Accuracy By Certainty Estimation of Human Decisions, Labels, and Raters](https://research.fb.com/blog/2020/08/improving-the-accuracy-of-community-standards-enforcement-by-certainty-estimation-of-human-decisions/) ([Paper](https://research.fb.com/wp-content/uploads/2020/08/CLARA-Confidence-of-Labels-and-Raters.pdf)) `Facebook` `2020`

## Data Engineering

1. [Zipline: Airbnb’s Machine Learning Data Management Platform](https://databricks.com/session/zipline-airbnbs-machine-learning-data-management-platform) `Airbnb` `2018`

2. [Sputnik: Airbnb’s Apache Spark Framework for Data Engineering](https://databricks.com/session\_na20/sputnik-airbnbs-apache-spark-framework-for-data-engineering) `Airbnb` `2020`

3. [Unbundling Data Science Workflows with Metaflow and AWS Step Functions](https://netflixtechblog.com/unbundling-data-science-workflows-with-metaflow-and-aws-step-functions-d454780c6280) `Netflix` `2020`

4. [How DoorDash is Scaling its Data Platform to Delight Customers and Meet Growing Demand](https://doordash.engineering/2020/09/25/how-doordash-is-scaling-its-data-platform/) `DoorDash` `2020`

5. [Revolutionizing Money Movements at Scale with Strong Data Consistency](https://eng.uber.com/money-scale-strong-data/) `Uber` `2020`

6. [Zipline - A Declarative Feature Engineering Framework](https://www.youtube.com/watch?v=LjcKCm0G\_OY) `Airbnb` `2020`

7. [Automating Data Protection at Scale, Part 1](https://medium.com/airbnb-engineering/automating-data-protection-at-scale-part-1-c74909328e08) ([Part 2](https://medium.com/airbnb-engineering/automating-data-protection-at-scale-part-2-c2b8d2068216)) `Airbnb` `2021`

8. [Real-time Data Infrastructure at Uber](https://arxiv.org/pdf/2104.00087.pdf) `Uber` `2021`

9. [Introducing Fabricator: A Declarative Feature Engineering Framework](https://doordash.engineering/2022/01/11/introducing-fabricator-a-declarative-feature-engineering-framework/) `DoorDash` `2022`

10. [Functions & DAGs: introducing Hamilton, a microframework for dataframe generation](https://multithreaded.stitchfix.com/blog/2021/10/14/functions-dags-hamilton/) `Stitch Fix` `2021`

11. [Optimizing Pinterest’s Data Ingestion Stack: Findings and Learnings](https://medium.com/@Pinterest\_Engineering/optimizing-pinterests-data-ingestion-stack-findings-and-learnings-994fddb063bf) `Pinterest` `2022`

12. [Lessons Learned From Running Apache Airflow at Scale](https://shopifyengineering.myshopify.com/blogs/engineering/lessons-learned-apache-airflow-scale) `Shopify` `2022`

13. [Understanding Data Storage and Ingestion for Large-Scale Deep Recommendation Model Training](https://arxiv.org/abs/2108.09373v4) `Meta` `2022`

## Data Discovery

1. [Apache Atlas: Data Goverance and Metadata Framework for Hadoop](https://atlas.apache.org/#/) ([Code](https://github.com/apache/atlas)) `Apache`

2. [Collect, Aggregate, and Visualize a Data Ecosystem's Metadata](https://marquezproject.github.io/marquez/) ([Code](https://github.com/MarquezProject/marquez)) `WeWork`

3. [Discovery and Consumption of Analytics Data at Twitter](https://blog.twitter.com/engineering/en\_us/topics/insights/2016/discovery-and-consumption-of-analytics-data-at-twitter.html) `Twitter` `2016`

4. [Democratizing Data at Airbnb](https://medium.com/airbnb-engineering/democratizing-data-at-airbnb-852d76c51770) `Airbnb` `2017`

5. [Databook: Turning Big Data into Knowledge with Metadata at Uber](https://eng.uber.com/databook/) `Uber` `2018`

6. [Metacat: Making Big Data Discoverable and Meaningful at Netflix](https://netflixtechblog.com/metacat-making-big-data-discoverable-and-meaningful-at-netflix-56fb36a53520) ([Code](https://github.com/Netflix/metacat)) `Netflix` `2018`

7. [Amundsen — Lyft’s Data Discovery & Metadata Engine](https://eng.lyft.com/amundsen-lyfts-data-discovery-metadata-engine-62d27254fbb9) `Lyft` `2019`

8. [Open Sourcing Amundsen: A Data Discovery And Metadata Platform](https://eng.lyft.com/open-sourcing-amundsen-a-data-discovery-and-metadata-platform-2282bb436234) ([Code](https://github.com/lyft/amundsen)) `Lyft` `2019`

9. [DataHub: A Generalized Metadata Search & Discovery Tool](https://engineering.linkedin.com/blog/2019/data-hub) ([Code](https://github.com/linkedin/datahub)) `LinkedIn` `2019`

10. [Amundsen: One Year Later](https://eng.lyft.com/amundsen-1-year-later-7b60bf28602) `Lyft` `2020`

11. [Using Amundsen to Support User Privacy via Metadata Collection at Square](https://developer.squareup.com/blog/using-amundsen-to-support-user-privacy-via-metadata-collection-at-square/) `Square` `2020`

12. [Turning Metadata Into Insights with Databook](https://eng.uber.com/metadata-insights-databook/) `Uber` `2020`

13. [DataHub: Popular Metadata Architectures Explained](https://engineering.linkedin.com/blog/2020/datahub-popular-metadata-architectures-explained) `LinkedIn` `2020`

14. [How We Improved Data Discovery for Data Scientists at Spotify](https://engineering.atspotify.com/2020/02/27/how-we-improved-data-discovery-for-data-scientists-at-spotify/) `Spotify` `2020`

15. [How We’re Solving Data Discovery Challenges at Shopify](https://engineering.shopify.com/blogs/engineering/solving-data-discovery-challenges-shopify) `Shopify` `2020`

16. [Nemo: Data discovery at Facebook](https://engineering.fb.com/data-infrastructure/nemo/) `Facebook` `2020`

17. [Exploring Data @ Netflix](https://netflixtechblog.com/exploring-data-netflix-9d87e20072e3) ([Code](https://github.com/Netflix/nf-data-explorer)) `Netflix` `2021`

## Feature Stores

1. [Distributed Time Travel for Feature Generation](https://netflixtechblog.com/distributed-time-travel-for-feature-generation-389cccdd3907) `Netflix` `2016`

2. [Building the Activity Graph, Part 2 (Feature Storage Section)](https://engineering.linkedin.com/blog/2017/07/building-the-activity-graph--part-2) `LinkedIn` `2017`

3. [Fact Store at Scale for Netflix Recommendations](https://databricks.com/session/fact-store-scale-for-netflix-recommendations) `Netflix` `2018`

4. [Zipline: Airbnb’s Machine Learning Data Management Platform](https://databricks.com/session/zipline-airbnbs-machine-learning-data-management-platform) `Airbnb` `2018`

5. [Introducing Feast: An Open Source Feature Store for Machine Learning](https://cloud.google.com/blog/products/ai-machine-learning/introducing-feast-an-open-source-feature-store-for-machine-learning) ([Code](https://github.com/feast-dev/feast)) `Gojek` `2019`

6. [Michelangelo Palette: A Feature Engineering Platform at Uber](https://www.infoq.com/presentations/michelangelo-palette-uber/) `Uber` `2019`

7. [The Architecture That Powers Twitter's Feature Store](https://www.youtube.com/watch?v=UNailXoiIrY) `Twitter` `2019`

8. [Accelerating Machine Learning with the Feature Store Service](https://technology.condenast.com/story/accelerating-machine-learning-with-the-feature-store-service) `Condé Nast` `2019`

9. [Feast: Bridging ML Models and Data](https://www.gojek.io/blog/feast-bridging-ml-models-and-data) `Gojek` `2020`

10. [Building a Scalable ML Feature Store with Redis, Binary Serialization, and Compression](https://doordash.engineering/2020/11/19/building-a-gigascale-ml-feature-store-with-redis/) `DoorDash` `2020`

11. [Rapid Experimentation Through Standardization: Typed AI features for LinkedIn’s Feed](https://engineering.linkedin.com/blog/2020/feed-typed-ai-features) `LinkedIn` `2020`

12. [Building a Feature Store](https://nlathia.github.io/2020/12/Building-a-feature-store.html) `Monzo Bank` `2020`

13. [Butterfree: A Spark-based Framework for Feature Store Building](https://medium.com/quintoandar-tech-blog/butterfree-a-spark-based-framework-for-feature-store-building-48c3640522c7) ([Code](https://github.com/quintoandar/butterfree)) `QuintoAndar` `2020`

14. [Building Riviera: A Declarative Real-Time Feature Engineering Framework](https://doordash.engineering/2021/03/04/building-a-declarative-real-time-feature-engineering-framework/) `DoorDash` `2021`

15. [Optimal Feature Discovery: Better, Leaner Machine Learning Models Through Information Theory](https://eng.uber.com/optimal-feature-discovery-ml/) `Uber` `2021`

16. [ML Feature Serving Infrastructure at Lyft](https://eng.lyft.com/ml-feature-serving-infrastructure-at-lyft-d30bf2d3c32a) `Lyft` `2021`

17. [Near real-time features for near real-time personalization](https://engineering.linkedin.com/blog/2022/near-real-time-features-for-near-real-time-personalization) `LinkedIn` `2022`

18. [Building the Model Behind DoorDash’s Expansive Merchant Selection](https://doordash.engineering/2022/04/19/building-merchant-selection/) `DoorDash` `2022`

19. [Open sourcing Feathr – LinkedIn’s feature store for productive machine learning](https://engineering.linkedin.com/blog/2022/open-sourcing-feathr---linkedin-s-feature-store-for-productive-m) `LinkedIn` `2022`

20. [Developing scalable feature engineering DAGs](https://outerbounds.com/blog/developing-scalable-feature-engineering-dags) `Metaflow + Hamilton` via `Outerbounds` `2022`

## Classification

1. [Prediction of Advertiser Churn for Google AdWords](https://research.google/pubs/pub36678/) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36678.pdf)) `Google` `2010`

2. [High-Precision Phrase-Based Document Classification on a Modern Scale](https://engineering.linkedin.com/research/2011/high-precision-phrase-based-document-classification-on-a-modern-scale) ([Paper](http://web.stanford.edu/~gavish/documents/phrase\_based.pdf)) `LinkedIn` `2011`

3. [Chimera: Large-scale Classification using Machine Learning, Rules, and Crowdsourcing](https://dl.acm.org/doi/10.14778/2733004.2733024) ([Paper](http://pages.cs.wisc.edu/%7Eanhai/papers/chimera-vldb14.pdf)) `Walmart` `2014`

4. [Large-scale Item Categorization in e-Commerce Using Multiple Recurrent Neural Networks](https://www.kdd.org/kdd2016/subtopic/view/large-scale-item-categorization-in-e-commerce-using-multiple-recurrent-neur/) ([Paper](https://www.kdd.org/kdd2016/papers/files/adf0392-haAemb.pdf)) `NAVER` `2016`

5. [Learning to Diagnose with LSTM Recurrent Neural Networks](https://arxiv.org/abs/1511.03677) ([Paper](https://arxiv.org/pdf/1511.03677.pdf)) `Google` `2017`

6. [Discovering and Classifying In-app Message Intent at Airbnb](https://medium.com/airbnb-engineering/discovering-and-classifying-in-app-message-intent-at-airbnb-6a55f5400a0c) `Airbnb` `2019`

7. [Teaching Machines to Triage Firefox Bugs](https://hacks.mozilla.org/2019/04/teaching-machines-to-triage-firefox-bugs/) `Mozilla` `2019`

8. [Categorizing Products at Scale](https://engineering.shopify.com/blogs/engineering/categorizing-products-at-scale) `Shopify` `2020`

9. [How We Built the Good First Issues Feature](https://github.blog/2020-01-22-how-we-built-good-first-issues/) `GitHub` `2020`

10. [Testing Firefox More Efficiently with Machine Learning](https://hacks.mozilla.org/2020/07/testing-firefox-more-efficiently-with-machine-learning/) `Mozilla` `2020`

11. [Using ML to Subtype Patients Receiving Digital Mental Health Interventions](https://www.microsoft.com/en-us/research/blog/a-path-to-personalization-using-ml-to-subtype-patients-receiving-digital-mental-health-interventions/) ([Paper](https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2768347)) `Microsoft` `2020`

12. [Scalable Data Classification for Security and Privacy](https://engineering.fb.com/security/data-classification-system/) ([Paper](https://arxiv.org/pdf/2006.14109.pdf)) `Facebook` `2020`

13. [Uncovering Online Delivery Menu Best Practices with Machine Learning](https://doordash.engineering/2020/11/10/uncovering-online-delivery-menu-best-practices-with-machine-learning/) `DoorDash` `2020`

14. [Using a Human-in-the-Loop to Overcome the Cold Start Problem in Menu Item Tagging](https://doordash.engineering/2020/08/28/overcome-the-cold-start-problem-in-menu-item-tagging/) `DoorDash` `2020`

15. [Deep Learning: Product Categorization and Shelving](https://medium.com/walmartglobaltech/deep-learning-product-categorization-and-shelving-630571e81e96) `Walmart` `2021`

16. [Large-scale Item Categorization for e-Commerce](https://dl.acm.org/doi/10.1145/2396761.2396838) ([Paper](https://www.researchgate.net/profile/Jean\_David\_Ruvini/publication/262270957\_Large-scale\_item\_categorization\_for\_e-commerce/links/5512dc3d0cf270fd7e33a0d5/Large-scale-item-categorization-for-e-commerce.pdf)) `DianPing`, `eBay` `2012`

17. [Semantic Label Representation with an Application on Multimodal Product Categorization](https://medium.com/walmartglobaltech/semantic-label-representation-with-an-application-on-multimodal-product-categorization-63d668b943b7) `Walmart` `2022`

## Regression

1. [Using Machine Learning to Predict Value of Homes On Airbnb](https://medium.com/airbnb-engineering/using-machine-learning-to-predict-value-of-homes-on-airbnb-9272d3d4739d) `Airbnb` `2017`

2. [Using Machine Learning to Predict the Value of Ad Requests](https://blog.twitter.com/engineering/en\_us/topics/insights/2020/using-machine-learning-to-predict-the-value-of-ad-requests.html) `Twitter` `2020`

3. [Open-Sourcing Riskquant, a Library for Quantifying Risk](https://netflixtechblog.com/open-sourcing-riskquant-a-library-for-quantifying-risk-6720cc1e4968) ([Code](https://github.com/Netflix-Skunkworks/riskquant)) `Netflix` `2020`

4. [Solving for Unobserved Data in a Regression Model Using a Simple Data Adjustment](https://doordash.engineering/2020/10/14/solving-for-unobserved-data-in-a-regression-model/) `DoorDash` `2020`

## Forecasting

1. [Engineering Extreme Event Forecasting at Uber with RNN](https://eng.uber.com/neural-networks/) `Uber` `2017`

2. [Forecasting at Uber: An Introduction](https://eng.uber.com/forecasting-introduction/) `Uber` `2018`

3. [Transforming Financial Forecasting with Data Science and Machine Learning at Uber](https://eng.uber.com/transforming-financial-forecasting-machine-learning/) `Uber` `2018`

4. [Under the Hood of Gojek’s Automated Forecasting Tool](https://www.gojek.io/blog/under-the-hood-of-gojeks-automated-forecasting-tool) `Gojek` `2019`

5. [BusTr: Predicting Bus Travel Times from Real-Time Traffic](https://dl.acm.org/doi/abs/10.1145/3394486.3403376) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3394486.3403376), [Video](https://crossminds.ai/video/5f3369790576dd25aef288db/)) `Google` `2020`

6. [Retraining Machine Learning Models in the Wake of COVID-19](https://doordash.engineering/2020/09/15/retraining-ml-models-covid-19/) `DoorDash` `2020`

7. [Automatic Forecasting using Prophet, Databricks, Delta Lake and MLflow](https://www.youtube.com/watch?v=TkcpjnLh690) ([Paper](https://peerj.com/preprints/3190.pdf), [Code](https://github.com/facebook/prophet)) `Atlassian` `2020`

8. [Introducing Orbit, An Open Source Package for Time Series Inference and Forecasting](https://eng.uber.com/orbit/) ([Paper](https://arxiv.org/abs/2004.08492), [Video](https://youtu.be/LXDpq\_iwcWY), [Code](https://github.com/uber/orbit)) `Uber` `2021`

9. [Managing Supply and Demand Balance Through Machine Learning](https://doordash.engineering/2021/06/29/managing-supply-and-demand-balance-through-machine-learning/) `DoorDash` `2021`

10. [Greykite: A flexible, intuitive, and fast forecasting library](https://engineering.linkedin.com/blog/2021/greykite--a-flexible--intuitive--and-fast-forecasting-library) `LinkedIn` `2021`

11. [DeepETA: How Uber Predicts Arrival Times Using Deep Learning](https://eng.uber.com/deepeta-how-uber-predicts-arrival-times/) `Uber` `2022`

12. [Forecasting Grubhub Order Volume At Scale](https://bytes.grubhub.com/forecasting-grubhub-order-volume-at-scale-a966c2f901d2) `Grubhub` `2022`

## Recommendation

1. [Amazon.com Recommendations: Item-to-Item Collaborative Filtering](https://ieeexplore.ieee.org/document/1167344) ([Paper](https://www.cs.umd.edu/~samir/498/Amazon-Recommendations.pdf)) `Amazon` `2003`

2. [Netflix Recommendations: Beyond the 5 stars (Part 1](https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429) ([Part 2](https://netflixtechblog.com/netflix-recommendations-beyond-the-5-stars-part-2-d9b96aa399f5)) `Netflix` `2012`

3. [How Music Recommendation Works — And Doesn’t Work](https://notes.variogr.am/2012/12/11/how-music-recommendation-works-and-doesnt-work/) `Spotify` `2012`

4. [Learning to Rank Recommendations with the k -Order Statistic Loss](https://dl.acm.org/doi/10.1145/2507157.2507210) ([Paper](https://dl.acm.org/doi/pdf/10.1145/2507157.2507210)) `Google` `2013`

5. [Recommending Music on Spotify with Deep Learning](https://benanne.github.io/2014/08/05/spotify-cnns.html) `Spotify` `2014`

6. [Learning a Personalized Homepage](https://netflixtechblog.com/learning-a-personalized-homepage-aa8ec670359a) `Netflix` `2015`

7. [Session-based Recommendations with Recurrent Neural Networks](https://arxiv.org/abs/1511.06939) ([Paper](https://arxiv.org/pdf/1511.06939.pdf)) `Telefonica` `2016`

8. [Deep Neural Networks for YouTube Recommendations](https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45530.pdf) `YouTube` `2016`

9. [E-commerce in Your Inbox: Product Recommendations at Scale](https://arxiv.org/abs/1606.07154) ([Paper](https://arxiv.org/pdf/1606.07154.pdf)) `Yahoo` `2016`

10. [To Be Continued: Helping you find shows to continue watching on Netflix](https://netflixtechblog.com/to-be-continued-helping-you-find-shows-to-continue-watching-on-7c0d8ee4dab6) `Netflix` `2016`

11. [Personalized Recommendations in LinkedIn Learning](https://engineering.linkedin.com/blog/2016/12/personalized-recommendations-in-linkedin-learning) `LinkedIn` `2016`

12. [Personalized Channel Recommendations in Slack](https://slack.engineering/personalized-channel-recommendations-in-slack/) `Slack` `2016`

13. [Recommending Complementary Products in E-Commerce Push Notifications](https://arxiv.org/abs/1707.08113) ([Paper](https://arxiv.org/pdf/1707.08113.pdf)) `Alibaba` `2017`

14. [Artwork Personalization at Netflix](https://netflixtechblog.com/artwork-personalization-c589f074ad76) `Netflix` `2017`

15. [A Meta-Learning Perspective on Cold-Start Recommendations for Items](https://papers.nips.cc/paper/7266-a-meta-learning-perspective-on-cold-start-recommendations-for-items) ([Paper](https://papers.nips.cc/paper/7266-a-meta-learning-perspective-on-cold-start-recommendations-for-items.pdf)) `Twitter` `2017`

16. [Pixie: A System for Recommending 3+ Billion Items to 200+ Million Users in Real-Time](https://arxiv.org/abs/1711.07601) ([Paper](https://arxiv.org/pdf/1711.07601.pdf)) `Pinterest` `2017`

17. [How 20th Century Fox uses ML to predict a movie audience](https://cloud.google.com/blog/products/ai-machine-learning/how-20th-century-fox-uses-ml-to-predict-a-movie-audience) ([Paper](https://arxiv.org/abs/1810.08189)) `20th Century Fox` `2018`

18. [Calibrated Recommendations](https://dl.acm.org/doi/10.1145/3240323.3240372) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3240323.3240372)) `Netflix` `2018`

19. [Food Discovery with Uber Eats: Recommending for the Marketplace](https://eng.uber.com/uber-eats-recommending-marketplace/) `Uber` `2018`

20. [Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits](https://dl.acm.org/doi/10.1145/3240323.3240354) ([Paper](https://static1.squarespace.com/static/5ae0d0b48ab7227d232c2bea/t/5ba849e3c83025fa56814f45/1537755637453/BartRecSys.pdf)) `Spotify` `2018`

21. [Behavior Sequence Transformer for E-commerce Recommendation in Alibaba](https://arxiv.org/abs/1905.06874) ([Paper](https://arxiv.org/pdf/1905.06874.pdf)) `Alibaba` `2019`

22. [SDM: Sequential Deep Matching Model for Online Large-scale Recommender System](https://arxiv.org/abs/1909.00385) ([Paper](https://arxiv.org/pdf/1909.00385.pdf)) `Alibaba` `2019`

23. [Multi-Interest Network with Dynamic Routing for Recommendation at Tmall](https://arxiv.org/abs/1904.08030) ([Paper](https://arxiv.org/pdf/1904.08030.pdf)) `Alibaba` `2019`

24. [Personalized Recommendations for Experiences Using Deep Learning](https://www.tripadvisor.com/engineering/personalized-recommendations-for-experiences-using-deep-learning/) `TripAdvisor` `2019`

25. [Powered by AI: Instagram’s Explore recommender system](https://ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/) `Facebook` `2019`

26. [Marginal Posterior Sampling for Slate Bandits](https://www.ijcai.org/proceedings/2019/308) ([Paper](https://www.ijcai.org/proceedings/2019/0308.pdf)) `Netflix` `2019`

27. [Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations](https://eng.uber.com/uber-eats-graph-learning/) `Uber` `2019`

28. [Music recommendation at Spotify](http://sigir.org/afirm2019/slides/16.%20Friday%20-%20Music%20Recommendation%20at%20Spotify%20-%20Ben%20Carterette.pdf) `Spotify` `2019`

29. [Using Machine Learning to Predict what File you Need Next (Part 1)](https://dropbox.tech/machine-learning/content-suggestions-machine-learning) `Dropbox` `2019`

30. [Using Machine Learning to Predict what File you Need Next (Part 2)](https://dropbox.tech/machine-learning/using-machine-learning-to-predict-what-file-you-need-next-part-2) `Dropbox` `2019`

31. [Learning to be Relevant: Evolution of a Course Recommendation System](https://dl.acm.org/doi/pdf/10.1145/3357384.3357817) (\*\*PAPER NEEDED\*\*)`LinkedIn` `2019`

32. [Temporal-Contextual Recommendation in Real-Time](https://www.amazon.science/publications/temporal-contextual-recommendation-in-real-time) ([Paper](https://assets.amazon.science/96/71/d1f25754497681133c7aa2b7eb05/temporal-contextual-recommendation-in-real-time.pdf)) `Amazon` `2020`

33. [P-Companion: A Framework for Diversified Complementary Product Recommendation](https://www.amazon.science/publications/p-companion-a-principled-framework-for-diversified-complementary-product-recommendation) ([Paper](https://assets.amazon.science/d5/16/3f7809974a899a11bacdadefdf24/p-companion-a-principled-framework-for-diversified-complementary-product-recommendation.pdf)) `Amazon` `2020`

34. [Deep Interest with Hierarchical Attention Network for Click-Through Rate Prediction](https://arxiv.org/abs/2005.12981) ([Paper](https://arxiv.org/pdf/2005.12981.pdf)) `Alibaba` `2020`

35. [TPG-DNN: A Method for User Intent Prediction with Multi-task Learning](https://arxiv.org/abs/2008.02122) ([Paper](https://arxiv.org/pdf/2008.02122.pdf)) `Alibaba` `2020`

36. [PURS: Personalized Unexpected Recommender System for Improving User Satisfaction](https://dl.acm.org/doi/10.1145/3383313.3412238) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3383313.3412238)) `Alibaba` `2020`

37. [Controllable Multi-Interest Framework for Recommendation](https://arxiv.org/abs/2005.09347) ([Paper](https://arxiv.org/pdf/2005.09347)) `Alibaba` `2020`

38. [MiNet: Mixed Interest Network for Cross-Domain Click-Through Rate Prediction](https://arxiv.org/abs/2008.02974) ([Paper](https://arxiv.org/pdf/2008.02974.pdf)) `Alibaba` `2020`

39. [ATBRG: Adaptive Target-Behavior Relational Graph Network for Effective Recommendation](https://arxiv.org/abs/2005.12002) ([Paper](https://arxiv.org/pdf/2005.12002.pdf)) `Alibaba` `2020`

40. [For Your Ears Only: Personalizing Spotify Home with Machine Learning](https://engineering.atspotify.com/2020/01/16/for-your-ears-only-personalizing-spotify-home-with-machine-learning/) `Spotify` `2020`

41. [Reach for the Top: How Spotify Built Shortcuts in Just Six Months](https://engineering.atspotify.com/2020/04/15/reach-for-the-top-how-spotify-built-shortcuts-in-just-six-months/) `Spotify` `2020`

42. [Contextual and Sequential User Embeddings for Large-Scale Music Recommendation](https://dl.acm.org/doi/10.1145/3383313.3412248) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3383313.3412248)) `Spotify` `2020`

43. [The Evolution of Kit: Automating Marketing Using Machine Learning](https://engineering.shopify.com/blogs/engineering/evolution-kit-automating-marketing-machine-learning) `Shopify` `2020`

44. [A Closer Look at the AI Behind Course Recommendations on LinkedIn Learning (Part 1)](https://engineering.linkedin.com/blog/2020/course-recommendations-ai-part-one) `LinkedIn` `2020`

45. [A Closer Look at the AI Behind Course Recommendations on LinkedIn Learning (Part 2)](https://engineering.linkedin.com/blog/2020/course-recommendations-ai-part-two) `LinkedIn` `2020`

46. [Building a Heterogeneous Social Network Recommendation System](https://engineering.linkedin.com/blog/2020/building-a-heterogeneous-social-network-recommendation-system) `LinkedIn` `2020`

47. [How TikTok recommends videos #ForYou](https://newsroom.tiktok.com/en-us/how-tiktok-recommends-videos-for-you) `ByteDance` `2020`

48. [Zero-Shot Heterogeneous Transfer Learning from RecSys to Cold-Start Search Retrieval](https://arxiv.org/abs/2008.02930) ([Paper](https://arxiv.org/pdf/2008.02930.pdf)) `Google` `2020`

49. [Improved Deep & Cross Network for Feature Cross Learning in Web-scale LTR Systems](https://arxiv.org/abs/2008.13535) ([Paper](https://arxiv.org/pdf/2008.13535.pdf)) `Google` `2020`

50. [Mixed Negative Sampling for Learning Two-tower Neural Networks in Recommendations](https://research.google/pubs/pub50257/) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/b9f4e78a8830fe5afcf2f0452862fb3c0d6584ea.pdf)) `Google` `2020`

51. [Future Data Helps Training: Modeling Future Contexts for Session-based Recommendation](https://arxiv.org/pdf/1906.04473.pdf) ([Paper](https://arxiv.org/pdf/1906.04473.pdf)) `Tencent` `2020`

52. [A Case Study of Session-based Recommendations in the Home-improvement Domain](https://dl.acm.org/doi/10.1145/3383313.3412235) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3383313.3412235)) `Home Depot` `2020`

53. [Balancing Relevance and Discovery to Inspire Customers in the IKEA App](https://dl.acm.org/doi/10.1145/3383313.3411550) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3383313.3411550)) `Ikea` `2020`

54. [How we use AutoML, Multi-task learning and Multi-tower models for Pinterest Ads](https://medium.com/pinterest-engineering/how-we-use-automl-multi-task-learning-and-multi-tower-models-for-pinterest-ads-db966c3dc99e) `Pinterest` `2020`

55. [Multi-task Learning for Related Products Recommendations at Pinterest](https://medium.com/pinterest-engineering/multi-task-learning-for-related-products-recommendations-at-pinterest-62684f631c12) `Pinterest` `2020`

56. [Improving the Quality of Recommended Pins with Lightweight Ranking](https://medium.com/pinterest-engineering/improving-the-quality-of-recommended-pins-with-lightweight-ranking-8ff5477b20e3) `Pinterest` `2020`

57. [Multi-task Learning and Calibration for Utility-based Home Feed Ranking](https://medium.com/pinterest-engineering/multi-task-learning-and-calibration-for-utility-based-home-feed-ranking-64087a7bcbad) `Pinterest` `2020`

57. [Personalized Cuisine Filter Based on Customer Preference and Local Popularity](https://doordash.engineering/2020/01/27/personalized-cuisine-filter/) `DoorDash` `2020`

58. [How We Built a Matchmaking Algorithm to Cross-Sell Products](https://www.gojek.io/blog/how-we-built-a-matchmaking-algorithm-to-cross-sell-products) `Gojek` `2020`

59. [Lessons Learned Addressing Dataset Bias in Model-Based Candidate Generation](https://arxiv.org/abs/2105.09293) ([Paper](https://arxiv.org/pdf/2105.09293.pdf)) `Twitter` `2021`

60. [Self-supervised Learning for Large-scale Item Recommendations](https://arxiv.org/abs/2007.12865) ([Paper](https://arxiv.org/pdf/2007.12865.pdf)) `Google` `2021`

61. [Deep Retrieval: End-to-End Learnable Structure Model for Large-Scale Recommendations](https://arxiv.org/abs/2007.07203) ([Paper](https://arxiv.org/pdf/2007.07203.pdf)) `ByteDance` `2021`

62. [Using AI to Help Health Experts Address the COVID-19 Pandemic](https://ai.facebook.com/blog/using-ai-to-help-health-experts-address-the-covid-19-pandemic/) `Facebook` `2021`

63. [Advertiser Recommendation Systems at Pinterest](https://medium.com/pinterest-engineering/advertiser-recommendation-systems-at-pinterest-ccb255fbde20) `Pinterest` `2021`

64. [On YouTube's Recommendation System](https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/) `YouTube` `2021`

65. [Mozrt, a Deep Learning Recommendation System Empowering Walmart Store Associates](https://medium.com/walmartglobaltech/mozrt-a-deep-learning-recommendation-system-empowering-walmart-store-associates-with-a-5d42c08d88da) `Walmart` `2021`

65. [The Amazon Music conversational recommender is hitting the right notes](https://www.amazon.science/latest-news/how-amazon-music-uses-recommendation-system-machine-learning) `Amazon` `2022`

66. [Personalized complementary product recommendation](https://www.amazon.science/publications/personalized-complementary-product-recommendation) ([Paper](https://assets.amazon.science/6c/d9/a0ec3eda4f0fb4312ce0ada41771/personalized-complementary-product-recommendation.pdf)) `Amazon` `2022`

67. [Building a Deep Learning Based Retrieval System for Personalized Recommendations](https://tech.ebayinc.com/engineering/building-a-deep-learning-based-retrieval-system-for-personalized-recommendations/) `eBay` `2022`

68. [How We Built: An Early-Stage Machine Learning Model for Recommendations](https://www.onepeloton.com/press/articles/how-we-built-machine-learning) `Peloton` `2022`

69. [Beyond Matrix Factorization: Using hybrid features for user-business recommendations](https://engineeringblog.yelp.com/2022/04/beyond-matrix-factorization-using-hybrid-features-for-user-business-recommendations.html) `Yelp` `2022`

70. [Improving job matching with machine-learned activity features](https://engineering.linkedin.com/blog/2022/improving-job-matching-with-machine-learned-activity-features-) `LinkedIn` `2022`

71. [Understanding Data Storage and Ingestion for Large-Scale Deep Recommendation Model Training](https://arxiv.org/abs/2108.09373v4) `Meta` `2022`

72. [Homepage Recommendation with Exploitation and Exploration](https://doordash.engineering/2022/10/05/homepage-recommendation-with-exploitation-and-exploration/) `DoorDash` `2022`

## Search & Ranking

1. [Amazon Search: The Joy of Ranking Products](https://www.amazon.science/publications/amazon-search-the-joy-of-ranking-products) ([Paper](https://assets.amazon.science/89/cd/34289f1f4d25b5857d776bdf04d5/amazon-search-the-joy-of-ranking-products.pdf), [Video](https://www.youtube.com/watch?v=NLrhmn-EZ88), [Code](https://github.com/dariasor/TreeExtra)) `Amazon` `2016`

2. [How Lazada Ranks Products to Improve Customer Experience and Conversion](https://www.slideshare.net/eugeneyan/how-lazada-ranks-products-to-improve-customer-experience-and-conversion) `Lazada` `2016`

3. [Ranking Relevance in Yahoo Search](https://www.kdd.org/kdd2016/subtopic/view/ranking-relevance-in-yahoo-search) ([Paper](https://www.kdd.org/kdd2016/papers/files/adf0361-yinA.pdf)) `Yahoo` `2016`

4. [Learning to Rank Personalized Search Results in Professional Networks](https://arxiv.org/abs/1605.04624) ([Paper](https://arxiv.org/pdf/1605.04624.pdf)) `LinkedIn` `2016`

5. [Using Deep Learning at Scale in Twitter’s Timelines](https://blog.twitter.com/engineering/en\_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html) `Twitter` `2017`

6. [An Ensemble-based Approach to Click-Through Rate Prediction for Promoted Listings at Etsy](https://arxiv.org/abs/1711.01377) ([Paper](https://arxiv.org/pdf/1711.01377.pdf)) `Etsy` `2017`

7. [Powering Search & Recommendations at DoorDash](https://doordash.engineering/2017/07/06/powering-search-recommendations-at-doordash/) `DoorDash` `2017`

8. [Applying Deep Learning To Airbnb Search](https://arxiv.org/abs/1810.09591) ([Paper](https://arxiv.org/pdf/1810.09591.pdf)) `Airbnb` `2018`

9. [In-session Personalization for Talent Search](https://arxiv.org/abs/1809.06488) ([Paper](https://arxiv.org/pdf/1809.06488.pdf)) `LinkedIn` `2018`

10. [Talent Search and Recommendation Systems at LinkedIn](https://arxiv.org/abs/1809.06481) ([Paper](https://arxiv.org/pdf/1809.06481.pdf)) `LinkedIn` `2018`

11. [Food Discovery with Uber Eats: Building a Query Understanding Engine](https://eng.uber.com/uber-eats-query-understanding/) `Uber` `2018`

12. [Globally Optimized Mutual Influence Aware Ranking in E-Commerce Search](https://arxiv.org/abs/1805.08524) ([Paper](https://arxiv.org/pdf/1805.08524.pdf)) `Alibaba` `2018`

13. [Reinforcement Learning to Rank in E-Commerce Search Engine](https://arxiv.org/abs/1803.00710) ([Paper](https://arxiv.org/pdf/1803.00710.pdf)) `Alibaba` `2018`

14. [Semantic Product Search](https://arxiv.org/abs/1907.00937) ([Paper](https://arxiv.org/pdf/1907.00937.pdf)) `Amazon` `2019`

15. [Machine Learning-Powered Search Ranking of Airbnb Experiences](https://medium.com/airbnb-engineering/machine-learning-powered-search-ranking-of-airbnb-experiences-110b4b1a0789) `Airbnb` `2019`

16. [Entity Personalized Talent Search Models with Tree Interaction Features](https://arxiv.org/abs/1902.09041) ([Paper](https://arxiv.org/pdf/1902.09041.pdf)) `LinkedIn` `2019`

17. [The AI Behind LinkedIn Recruiter Search and recommendation systems](https://engineering.linkedin.com/blog/2019/04/ai-behind-linkedin-recruiter-search-and-recommendation-systems) `LinkedIn` `2019`

18. [Learning Hiring Preferences: The AI Behind LinkedIn Jobs](https://engineering.linkedin.com/blog/2019/02/learning-hiring-preferences--the-ai-behind-linkedin-jobs) `LinkedIn` `2019`

19. [The Secret Sauce Behind Search Personalisation](https://www.gojek.io/blog/the-secret-sauce-behind-search-personalisation) `Gojek` `2019`

20. [Neural Code Search: ML-based Code Search Using Natural Language Queries](https://ai.facebook.com/blog/neural-code-search-ml-based-code-search-using-natural-language-queries/) `Facebook` `2019`

21. [Aggregating Search Results from Heterogeneous Sources via Reinforcement Learning](https://arxiv.org/abs/1902.08882) ([Paper](https://arxiv.org/pdf/1902.08882.pdf)) `Alibaba` `2019`

22. [Cross-domain Attention Network with Wasserstein Regularizers for E-commerce Search](https://dl.acm.org/doi/10.1145/3357384.3357809) `Alibaba` `2019`

23. [Understanding Searches Better Than Ever Before](https://www.blog.google/products/search/search-language-understanding-bert/) ([Paper](https://arxiv.org/pdf/1810.04805.pdf)) `Google` `2019`

24. [How We Used Semantic Search to Make Our Search 10x Smarter](https://medium.com/tokopedia-engineering/how-we-used-semantic-search-to-make-our-search-10x-smarter-bd9c7f601821) `Tokopedia` `2019`

25. [Query2vec: Search query expansion with query embeddings](https://bytes.grubhub.com/search-query-embeddings-using-query2vec-f5931df27d79) `GrubHub` `2019`

26. [MOBIUS: Towards the Next Generation of Query-Ad Matching in Baidu’s Sponsored Search](http://research.baidu.com/Public/uploads/5d12eca098d40.pdf) `Baidu` `2019`

27. [Why Do People Buy Seemingly Irrelevant Items in Voice Product Search?](https://www.amazon.science/publications/why-do-people-buy-irrelevant-items-in-voice-product-search) ([Paper](https://assets.amazon.science/f7/48/0562b2c14338a0b76ccf4f523fa5/why-do-people-buy-irrelevant-items-in-voice-product-search.pdf)) `Amazon` `2020`

28. [Managing Diversity in Airbnb Search](https://arxiv.org/abs/2004.02621) ([Paper](https://arxiv.org/pdf/2004.02621.pdf)) `Airbnb` `2020`

29. [Improving Deep Learning for Airbnb Search](https://arxiv.org/abs/2002.05515) ([Paper](https://arxiv.org/pdf/2002.05515.pdf)) `Airbnb` `2020`

30. [Quality Matches Via Personalized AI for Hirer and Seeker Preferences](https://engineering.linkedin.com/blog/2020/quality-matches-via-personalized-ai) `LinkedIn` `2020`

31. [Understanding Dwell Time to Improve LinkedIn Feed Ranking](https://engineering.linkedin.com/blog/2020/understanding-feed-dwell-time) `LinkedIn` `2020`

32. [Ads Allocation in Feed via Constrained Optimization](https://dl.acm.org/doi/abs/10.1145/3394486.3403391) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3394486.3403391), [Video](https://crossminds.ai/video/5f33697a0576dd25aef288ea/)) `LinkedIn` `2020`

33. [Understanding Dwell Time to Improve LinkedIn Feed Ranking](https://engineering.linkedin.com/blog/2020/understanding-feed-dwell-time) `LinkedIn` `2020`

34. [AI at Scale in Bing](https://blogs.bing.com/search/2020\_05/AI-at-Scale-in-Bing) `Microsoft` `2020`

35. [Query Understanding Engine in Traveloka Universal Search](https://medium.com/traveloka-engineering/query-understanding-engine-in-traveloka-universal-search-410ad3895db7) `Traveloka` `2020`

36. [Bayesian Product Ranking at Wayfair](https://tech.wayfair.com/data-science/2020/01/bayesian-product-ranking-at-wayfair) `Wayfair` `2020`

37. [COLD: Towards the Next Generation of Pre-Ranking System](https://arxiv.org/abs/2007.16122) ([Paper](https://arxiv.org/pdf/2007.16122.pdf)) `Alibaba` `2020`

38. [Shop The Look: Building a Large Scale Visual Shopping System at Pinterest](https://dl.acm.org/doi/abs/10.1145/3394486.3403372) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3394486.3403372), [Video](https://crossminds.ai/video/5f3369790576dd25aef288d7/)) `Pinterest` `2020`

39. [Driving Shopping Upsells from Pinterest Search](https://medium.com/pinterest-engineering/driving-shopping-upsells-from-pinterest-search-d06329255402) `Pinterest` `2020`

40. [GDMix: A Deep Ranking Personalization Framework](https://engineering.linkedin.com/blog/2020/gdmix--a-deep-ranking-personalization-framework) ([Code](https://github.com/linkedin/gdmix)) `LinkedIn` `2020`

41. [Bringing Personalized Search to Etsy](https://codeascraft.com/2020/10/29/bringing-personalized-search-to-etsy/) `Etsy` `2020`

42. [Building a Better Search Engine for Semantic Scholar](https://medium.com/ai2-blog/building-a-better-search-engine-for-semantic-scholar-ea23a0b661e7) `Allen Institute for AI` `2020`

43. [Query Understanding for Natural Language Enterprise Search](https://arxiv.org/abs/2012.06238) ([Paper](https://arxiv.org/pdf/2012.06238.pdf)) `Salesforce` `2020`

44. [Things Not Strings: Understanding Search Intent with Better Recall](https://doordash.engineering/2020/12/15/understanding-search-intent-with-better-recall/) `DoorDash` `2020`

45. [Query Understanding for Surfacing Under-served Music Content](https://research.atspotify.com/publications/query-understanding-for-surfacing-under-served-music-content/) ([Paper](https://labtomarket.files.wordpress.com/2020/08/cikm2020.pdf)) `Spotify` `2020`

46. [Embedding-based Retrieval in Facebook Search](https://arxiv.org/abs/2006.11632) ([Paper](https://arxiv.org/pdf/2006.11632.pdf)) `Facebook` `2020`

47. [Towards Personalized and Semantic Retrieval for E-commerce Search via Embedding Learning](https://arxiv.org/abs/2006.02282) ([Paper](https://arxiv.org/pdf/2006.02282.pdf)) `JD` `2020`

48. [QUEEN: Neural query rewriting in e-commerce](https://www.amazon.science/publications/queen-neural-query-rewriting-in-e-commerce) ([Paper](https://assets.amazon.science/f9/78/dda8f1e143dba8ca96e43ec487c6/queen-neural-query-rewriting-in-ecommerce.pdf)) `Amazon` `2021`

49. [Using Learning-to-rank to Precisely Locate Where to Deliver Packages](https://www.amazon.science/blog/using-learning-to-rank-to-precisely-locate-where-to-deliver-packages) ([Paper](https://www.amazon.science/publications/getting-your-package-to-the-right-place-supervised-machine-learning-for-geolocation)) `Amazon` `2021`

50. [Seasonal relevance in e-commerce search](https://www.amazon.science/publications/seasonal-relevance-in-e-commerce-search) ([Paper](https://assets.amazon.science/ac/5e/d47612a846d6bec15738d7c8ab40/seasonal-relevance-in-ecommerce-search.pdf)) `Amazon` `2021`

51. [Graph Intention Network for Click-through Rate Prediction in Sponsored Search](https://arxiv.org/abs/2103.16164) ([Paper](https://arxiv.org/pdf/2103.16164.pdf)) `Alibaba` `2021`

52. [How We Built A Context-Specific Bidding System for Etsy Ads](https://codeascraft.com/2021/03/23/how-we-built-a-context-specific-bidding-system-for-etsy-ads/) `Etsy` `2021`

53. [Pre-trained Language Model based Ranking in Baidu Search](https://arxiv.org/abs/2105.11108) ([Paper](https://arxiv.org/pdf/2105.11108.pdf)) `Baidu` `2021`

54. [Stitching together spaces for query-based recommendations](https://multithreaded.stitchfix.com/blog/2021/08/13/stitching-together-spaces-for-query-based-recommendations/) `Stitch Fix` `2021`

55. [Deep Natural Language Processing for LinkedIn Search Systems](https://arxiv.org/abs/2108.08252) ([Paper](https://arxiv.org/pdf/2108.08252.pdf)) `LinkedIn` `2021`

56. [Siamese BERT-based Model for Web Search Relevance Ranking](https://arxiv.org/abs/2112.01810) ([Paper](https://arxiv.org/pdf/2112.01810.pdf), [Code](https://github.com/seznam/DaReCzech)) `Seznam` `2021`

57. [SearchSage: Learning Search Query Representations at Pinterest](https://medium.com/pinterest-engineering/searchsage-learning-search-query-representations-at-pinterest-654f2bb887fc) `Pinterest` `2021`

58. [3 Changes to Expand DoorDash’s Product Search Beyond Delivery](https://doordash.engineering/2022/05/10/3-changes-to-expand-doordashs-product-search/) `DoorDash` `2022`

## Embeddings

1. [Vector Representation Of Items, Customer And Cart To Build A Recommendation System](https://arxiv.org/abs/1705.06338) ([Paper](https://arxiv.org/pdf/1705.06338.pdf)) `Sears` `2017`

2. [Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba](https://arxiv.org/abs/1803.02349) ([Paper](https://arxiv.org/pdf/1803.02349.pdf)) `Alibaba` `2018`

3. [Embeddings@Twitter](https://blog.twitter.com/engineering/en\_us/topics/insights/2018/embeddingsattwitter.html) `Twitter` `2018`

4. [Listing Embeddings in Search Ranking](https://medium.com/airbnb-engineering/listing-embeddings-for-similar-listing-recommendations-and-real-time-personalization-in-search-601172f7603e) ([Paper](https://www.kdd.org/kdd2018/accepted-papers/view/real-time-personalization-using-embeddings-for-search-ranking-at-airbnb)) `Airbnb` `2018`

5. [Understanding Latent Style](https://multithreaded.stitchfix.com/blog/2018/06/28/latent-style/) `Stitch Fix` `2018`

6. [Towards Deep and Representation Learning for Talent Search at LinkedIn](https://arxiv.org/abs/1809.06473) ([Paper](https://arxiv.org/pdf/1809.06473.pdf)) `LinkedIn` `2018`

7. [Personalized Store Feed with Vector Embeddings](https://doordash.engineering/2018/04/02/personalized-store-feed-with-vector-embeddings/) `DoorDash` `2018`

8. [Should we Embed? A Study on Performance of Embeddings for Real-Time Recommendations](https://arxiv.org/abs/1907.06556)([Paper](https://arxiv.org/pdf/1907.06556.pdf)) `Moshbit` `2019`

9. [Machine Learning for a Better Developer Experience](https://netflixtechblog.com/machine-learning-for-a-better-developer-experience-1e600c69f36c) `Netflix` `2020`

10. [Announcing ScaNN: Efficient Vector Similarity Search](https://ai.googleblog.com/2020/07/announcing-scann-efficient-vector.html) ([Paper](https://arxiv.org/pdf/1908.10396.pdf), [Code](https://github.com/google-research/google-research/tree/master/scann)) `Google` `2020`

11. [Embedding-based Retrieval at Scribd](https://tech.scribd.com/blog/2021/embedding-based-retrieval-scribd.html) `Scribd` `2021`

## Natural Language Processing

1. [Abusive Language Detection in Online User Content](https://dl.acm.org/doi/10.1145/2872427.2883062) ([Paper](http://www.yichang-cs.com/yahoo/WWW16\_Abusivedetection.pdf)) `Yahoo` `2016`

2. [Smart Reply: Automated Response Suggestion for Email](https://research.google/pubs/pub45189/) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/45189.pdf)) `Google` `2016`

3. [Building Smart Replies for Member Messages](https://engineering.linkedin.com/blog/2017/10/building-smart-replies-for-member-messages) `LinkedIn` `2017`

4. [Active Annotation: bootstrapping annotation lexicon and guidelines for supervised NLU learning](https://arxiv.org/abs/1908.04092) ([Paper](https://www.isca-speech.org/archive\_v0/Interspeech\_2019/pdfs/2537.pdf)) `Trento University` `2019`

5. [How Natural Language Processing Helps LinkedIn Members Get Support Easily](https://engineering.linkedin.com/blog/2019/04/how-natural-language-processing-help-support) `LinkedIn` `2019`

6. [Gmail Smart Compose: Real-Time Assisted Writing](https://arxiv.org/abs/1906.00080) ([Paper](https://arxiv.org/pdf/1906.00080.pdf)) `Google` `2019`

7. [Goal-Oriented End-to-End Conversational Models with Profile Features in a Real-World Setting](https://www.amazon.science/publications/goal-oriented-end-to-end-chatbots-with-profile-features-in-a-real-world-setting) ([Paper](https://assets.amazon.science/47/03/e0d14dc34d3eb6e0d4ec282067bd/goal-oriented-end-to-end-chatbots-with-profile-features-in-a-real-world-setting.pdf)) `Amazon` `2019`

8. [Give Me Jeans not Shoes: How BERT Helps Us Deliver What Clients Want](https://multithreaded.stitchfix.com/blog/2019/07/15/give-me-jeans/) `Stitch Fix` `2019`

9. [DeText: A deep NLP Framework for Intelligent Text Understanding](https://engineering.linkedin.com/blog/2020/open-sourcing-detext) ([Code](https://github.com/linkedin/detext)) `LinkedIn` `2020`

10. [SmartReply for YouTube Creators](https://ai.googleblog.com/2020/07/smartreply-for-youtube-creators.html) `Google` `2020`

11. [Using Neural Networks to Find Answers in Tables](https://ai.googleblog.com/2020/04/using-neural-networks-to-find-answers.html) ([Paper](https://arxiv.org/pdf/2004.02349.pdf)) `Google` `2020`

12. [A Scalable Approach to Reducing Gender Bias in Google Translate](https://ai.googleblog.com/2020/04/a-scalable-approach-to-reducing-gender.html) `Google` `2020`

13. [Assistive AI Makes Replying Easier](https://www.microsoft.com/en-us/research/group/msai/articles/assistive-ai-makes-replying-easier-2/) `Microsoft` `2020`

14. [AI Advances to Better Detect Hate Speech](https://ai.facebook.com/blog/ai-advances-to-better-detect-hate-speech/) `Facebook` `2020`

15. [A State-of-the-Art Open Source Chatbot](https://ai.facebook.com/blog/state-of-the-art-open-source-chatbot) ([Paper](https://arxiv.org/pdf/2004.13637.pdf)) `Facebook` `2020`

16. [A Highly Efficient, Real-Time Text-to-Speech System Deployed on CPUs](https://ai.facebook.com/blog/a-highly-efficient-real-time-text-to-speech-system-deployed-on-cpus/) `Facebook` `2020`

17. [Deep Learning to Translate Between Programming Languages](https://ai.facebook.com/blog/deep-learning-to-translate-between-programming-languages/) ([Paper](https://arxiv.org/abs/2006.03511), [Code](https://github.com/facebookresearch/TransCoder)) `Facebook` `2020`

18. [Deploying Lifelong Open-Domain Dialogue Learning](https://arxiv.org/abs/2008.08076) ([Paper](https://arxiv.org/pdf/2008.08076.pdf)) `Facebook` `2020`

19. [Introducing Dynabench: Rethinking the way we benchmark AI](https://ai.facebook.com/blog/dynabench-rethinking-ai-benchmarking/) `Facebook` `2020`

20. [How Gojek Uses NLP to Name Pickup Locations at Scale](https://www.gojek.io/blog/nlp-cartobert) `Gojek` `2020`

21. [The State-of-the-art Open-Domain Chatbot in Chinese and English](http://research.baidu.com/Blog/index-view?id=142) ([Paper](https://arxiv.org/pdf/2006.16779.pdf)) `Baidu` `2020`

22. [PEGASUS: A State-of-the-Art Model for Abstractive Text Summarization](https://ai.googleblog.com/2020/06/pegasus-state-of-art-model-for.html) ([Paper](https://arxiv.org/pdf/1912.08777.pdf), [Code](https://github.com/google-research/pegasus)) `Google` `2020`

23. [Photon: A Robust Cross-Domain Text-to-SQL System](https://www.aclweb.org/anthology/2020.acl-demos.24/) ([Paper](https://www.aclweb.org/anthology/2020.acl-demos.24.pdf)) ([Demo](http://naturalsql.com)) `Salesforce` `2020`

24. [GeDi: A Powerful New Method for Controlling Language Models](https://blog.einstein.ai/gedi/) ([Paper](https://arxiv.org/abs/2009.06367), [Code](https://github.com/salesforce/GeDi)) `Salesforce` `2020`

25. [Applying Topic Modeling to Improve Call Center Operations](https://www.youtube.com/watch?v=kzRR8OjF\_eI&t=2s) `RICOH` `2020`

26. [WIDeText: A Multimodal Deep Learning Framework](https://medium.com/airbnb-engineering/widetext-a-multimodal-deep-learning-framework-31ce2565880c) `Airbnb` `2020`

27. [Dynaboard: Moving Beyond Accuracy to Holistic Model Evaluation in NLP](https://ai.facebook.com/blog/dynaboard-moving-beyond-accuracy-to-holistic-model-evaluation-in-nlp) ([Code](https://github.com/facebookresearch/dynalab?fbclid=IwAR3qcV7QK2uXm4s4M0XUoQQo4i2DEsDy0LZFKxSQCHhP-3hF6fr2-NDFWX8)) `Facebook` `2021`

28. [How we reduced our text similarity runtime by 99.96%](https://medium.com/data-science-at-microsoft/how-we-reduced-our-text-similarity-runtime-by-99-96-e8e4b4426b35) `Microsoft` `2021`

29. [Textless NLP: Generating expressive speech from raw audio](https://ai.facebook.com/blog/textless-nlp-generating-expressive-speech-from-raw-audio/) [(Part 1)](https://arxiv.org/abs/2102.01192) [(Part 2)](https://arxiv.org/abs/2104.00355) [(Part 3)](https://arxiv.org/abs/2109.03264) [(Code and Pretrained Models)](https://github.com/pytorch/fairseq/tree/master/examples/textless\_nlp) `Facebook` `2021`

30. [Grammar Correction as You Type, on Pixel 6](https://ai.googleblog.com/2021/10/grammar-correction-as-you-type-on-pixel.html) `Google` `2021`

31. [Auto-generated Summaries in Google Docs](https://ai.googleblog.com/2022/03/auto-generated-summaries-in-google-docs.html) `Google` `2022`

32. [Words All the Way Down — Conversational Sentiment Analysis](https://medium.com/paypal-tech/words-all-the-way-down-conversational-sentiment-analysis-afe0165b84db) `PayPal` `2022`

## Sequence Modelling

1. [Doctor AI: Predicting Clinical Events via Recurrent Neural Networks](https://arxiv.org/abs/1511.05942) ([Paper](https://arxiv.org/pdf/1511.05942.pdf)) `Sutter Health` `2015`

2. [Deep Learning for Understanding Consumer Histories](https://engineering.zalando.com/posts/2016/10/deep-learning-for-understanding-consumer-histories.html) ([Paper](https://doogkong.github.io/2017/papers/paper2.pdf)) `Zalando` `2016`

3. [Using Recurrent Neural Network Models for Early Detection of Heart Failure Onset](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5391725/) ([Paper](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5391725/pdf/ocw112.pdf)) `Sutter Health` `2016`

4. [Continual Prediction of Notification Attendance with Classical and Deep Networks](https://arxiv.org/abs/1712.07120) ([Paper](https://arxiv.org/pdf/1712.07120.pdf)) `Telefonica` `2017`

5. [Deep Learning for Electronic Health Records](https://ai.googleblog.com/2018/05/deep-learning-for-electronic-health.html) ([Paper](https://www.nature.com/articles/s41746-018-0029-1.pdf)) `Google` `2018`

6. [Practice on Long Sequential User Behavior Modeling for Click-Through Rate Prediction](https://arxiv.org/abs/1905.09248) ([Paper](https://arxiv.org/pdf/1905.09248.pdf))`Alibaba` `2019`

7. [Search-based User Interest Modeling with Sequential Behavior Data for CTR Prediction](https://arxiv.org/abs/2006.05639) ([Paper](https://arxiv.org/pdf/2006.05639.pdf)) `Alibaba` `2020`

8. [How Duolingo uses AI in every part of its app](https://venturebeat.com/2020/08/18/how-duolingo-uses-ai-in-every-part-of-its-app/) `Duolingo` `2020`

9. [Leveraging Online Social Interactions For Enhancing Integrity at Facebook](https://research.fb.com/blog/2020/08/leveraging-online-social-interactions-for-enhancing-integrity-at-facebook/) ([Paper](https://research.fb.com/wp-content/uploads/2020/08/TIES-Temporal-Interaction-Embeddings-For-Enhancing-Social-Media-Integrity-At-Facebook.pdf), [Video](https://crossminds.ai/video/5f3369780576dd25aef288cf/)) `Facebook` `2020`

10. [Using deep learning to detect abusive sequences of member activity](https://engineering.linkedin.com/blog/2021/using-deep-learning-to-detect-abusive-sequences-of-member-activi) ([Video](https://exchange.scale.com/public/videos/using-deep-learning-to-detect-abusive-sequences-of-member-activity-on-linkedin)) `LinkedIn` `2021`

## Computer Vision

1. [Creating a Modern OCR Pipeline Using Computer Vision and Deep Learning](https://dropbox.tech/machine-learning/creating-a-modern-ocr-pipeline-using-computer-vision-and-deep-learning) `Dropbox` `2017`

2. [Categorizing Listing Photos at Airbnb](https://medium.com/airbnb-engineering/categorizing-listing-photos-at-airbnb-f9483f3ab7e3) `Airbnb` `2018`

3. [Amenity Detection and Beyond — New Frontiers of Computer Vision at Airbnb](https://medium.com/airbnb-engineering/amenity-detection-and-beyond-new-frontiers-of-computer-vision-at-airbnb-144a4441b72e) `Airbnb` `2019`

4. [How we Improved Computer Vision Metrics by More Than 5% Only by Cleaning Labelling Errors](https://deepomatic.com/en/how-we-improved-computer-vision-metrics-by-more-than-5-percent-only-by-cleaning-labelling-errors/) `Deepomatic`

5. [Making machines recognize and transcribe conversations in meetings using audio and video](https://www.microsoft.com/en-us/research/blog/making-machines-recognize-and-transcribe-conversations-in-meetings-using-audio-and-video/) `Microsoft` `2019`

6. [Powered by AI: Advancing product understanding and building new shopping experiences](https://ai.facebook.com/blog/powered-by-ai-advancing-product-understanding-and-building-new-shopping-experiences/) `Facebook` `2020`

7. [A Neural Weather Model for Eight-Hour Precipitation Forecasting](https://ai.googleblog.com/2020/03/a-neural-weather-model-for-eight-hour.html) ([Paper](https://arxiv.org/pdf/2003.12140.pdf)) `Google` `2020`

8. [Machine Learning-based Damage Assessment for Disaster Relief](https://ai.googleblog.com/2020/06/machine-learning-based-damage.html) ([Paper](https://arxiv.org/pdf/1910.06444.pdf)) `Google` `2020`

9. [RepNet: Counting Repetitions in Videos](https://ai.googleblog.com/2020/06/repnet-counting-repetitions-in-videos.html) ([Paper](https://openaccess.thecvf.com/content\_CVPR\_2020/papers/Dwibedi\_Counting\_Out\_Time\_Class\_Agnostic\_Video\_Repetition\_Counting\_in\_the\_CVPR\_2020\_paper.pdf)) `Google` `2020`

10. [Converting Text to Images for Product Discovery](https://www.amazon.science/blog/converting-text-to-images-for-product-discovery) ([Paper](https://assets.amazon.science/4c/76/5830542547b7a11089ce3af943b4/scipub-972.pdf)) `Amazon` `2020`

11. [How Disney Uses PyTorch for Animated Character Recognition](https://medium.com/pytorch/how-disney-uses-pytorch-for-animated-character-recognition-a1722a182627) `Disney` `2020`

12. [Image Captioning as an Assistive Technology](https://www.ibm.com/blogs/research/2020/07/image-captioning-assistive-technology/) ([Video](https://ivc.ischool.utexas.edu/~yz9244/VizWiz\_workshop/videos/MMTeam-oral.mp4)) `IBM` `2020`

13. [AI for AG: Production machine learning for agriculture](https://medium.com/pytorch/ai-for-ag-production-machine-learning-for-agriculture-e8cfdb9849a1) `Blue River` `2020`

14. [AI for Full-Self Driving at Tesla](https://youtu.be/hx7BXih7zx8?t=513) `Tesla` `2020`

15. [On-device Supermarket Product Recognition](https://ai.googleblog.com/2020/07/on-device-supermarket-product.html) `Google` `2020`

16. [Using Machine Learning to Detect Deficient Coverage in Colonoscopy Screenings](https://ai.googleblog.com/2020/08/using-machine-learning-to-detect.html) ([Paper](https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9097918)) `Google` `2020`

17. [Shop The Look: Building a Large Scale Visual Shopping System at Pinterest](https://dl.acm.org/doi/abs/10.1145/3394486.3403372) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3394486.3403372), [Video](https://crossminds.ai/video/5f3369790576dd25aef288d7/)) `Pinterest` `2020`

18. [Developing Real-Time, Automatic Sign Language Detection for Video Conferencing](https://ai.googleblog.com/2020/10/developing-real-time-automatic-sign.html) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/2eaf0d18ec6bef00d7dd88f39dd4f9ff13eeeeb2.pdf)) `Google` `2020`

19. [Vision-based Price Suggestion for Online Second-hand Items](https://arxiv.org/abs/2012.06009) ([Paper](https://arxiv.org/pdf/2012.06009.pdf)) `Alibaba` `2020`

20. [New AI Research to Help Predict COVID-19 Resource Needs From X-rays](https://ai.facebook.com/blog/new-ai-research-to-help-predict-covid-19-resource-needs-from-a-series-of-x-rays/) ([Paper](https://arxiv.org/pdf/2101.04909.pdf), [Model](https://github.com/facebookresearch/CovidPrognosis)) `Facebook` `2021`

21. [An Efficient Training Approach for Very Large Scale Face Recognition](https://arxiv.org/abs/2105.10375) ([Paper](https://arxiv.org/pdf/2105.10375)) `Alibaba` `2021`

22. [Identifying Document Types at Scribd](https://tech.scribd.com/blog/2021/identifying-document-types.html) `Scribd` `2021`

23. [Semi-Supervised Visual Representation Learning for Fashion Compatibility](https://arxiv.org/pdf/2109.08052.pdf) ([Paper](https://arxiv.org/pdf/2109.08052.pdf)) `Walmart` `2021`

24. [Recognizing People in Photos Through Private On-Device Machine Learning](https://machinelearning.apple.com/research/recognizing-people-photos) `Apple` `2021`

25. [DeepFusion: Lidar-Camera Deep Fusion for Multi-Modal 3D Object Detection](https://arxiv.org/pdf/2203.08195.pdf) `Google` `2022`

## Reinforcement Learning

1. [Deep Reinforcement Learning for Sponsored Search Real-time Bidding](https://arxiv.org/abs/1803.00259) ([Paper](https://arxiv.org/pdf/1803.00259.pdf)) `Alibaba` `2018`

2. [Budget Constrained Bidding by Model-free Reinforcement Learning in Display Advertising](https://arxiv.org/abs/1802.08365) ([Paper](https://arxiv.org/pdf/1802.08365.pdf)) `Alibaba` `2018`

3. [Reinforcement Learning for On-Demand Logistics](https://doordash.engineering/2018/09/10/reinforcement-learning-for-on-demand-logistics/) `DoorDash` `2018`

4. [Reinforcement Learning to Rank in E-Commerce Search Engine](https://arxiv.org/abs/1803.00710) ([Paper](https://arxiv.org/pdf/1803.00710.pdf)) `Alibaba` `2018`

5. [Dynamic Pricing on E-commerce Platform with Deep Reinforcement Learning](https://arxiv.org/abs/1912.02572) ([Paper](https://arxiv.org/pdf/1912.02572.pdf)) `Alibaba` `2019`

6. [Productionizing Deep Reinforcement Learning with Spark and MLflow](https://databricks.com/session\_na20/productionizing-deep-reinforcement-learning-with-spark-and-mlflow) `Zynga` `2020`

7. [Deep Reinforcement Learning in Production Part1](https://towardsdatascience.com/deep-reinforcement-learning-in-production-7e1e63471e2) [Part 2](https://towardsdatascience.com/deep-reinforcement-learning-in-production-part-2-personalizing-user-notifications-812a68ce2355) `Zynga` `2020`

8. [Building AI Trading Systems](https://dennybritz.com/blog/ai-trading/) `Denny Britz` `2020`

9. [Shifting Consumption towards Diverse content via Reinforcement Learning](https://research.atspotify.com/shifting-consumption-towards-diverse-content-via-reinforcement-learning/) ([Paper](https://dl.acm.org/doi/10.1145/3437963.3441775)) `Spotify` `2022`

## Anomaly Detection

1. [Detecting Performance Anomalies in External Firmware Deployments](https://netflixtechblog.com/detecting-performance-anomalies-in-external-firmware-deployments-ed41b1bfcf46) `Netflix` `2019`

2. [Detecting and Preventing Abuse on LinkedIn using Isolation Forests](https://engineering.linkedin.com/blog/2019/isolation-forest) ([Code](https://github.com/linkedin/isolation-forest)) `LinkedIn` `2019`

3. [Deep Anomaly Detection with Spark and Tensorflow](https://databricks.com/session\_eu19/deep-anomaly-detection-from-research-to-production-leveraging-spark-and-tensorflow) [(Hopsworks Video](https://www.youtube.com/watch?v=TgXVU8DSyCQ)) `Swedbank`, `Hopsworks` `2019`

4. [Preventing Abuse Using Unsupervised Learning](https://databricks.com/session\_na20/preventing-abuse-using-unsupervised-learning) `LinkedIn` `2020`

5. [The Technology Behind Fighting Harassment on LinkedIn](https://engineering.linkedin.com/blog/2020/fighting-harassment) `LinkedIn` `2020`

6. [Uncovering Insurance Fraud Conspiracy with Network Learning](https://arxiv.org/abs/2002.12789) ([Paper](https://arxiv.org/pdf/2002.12789.pdf)) `Ant Financial` `2020`

7. [How Does Spam Protection Work on Stack Exchange?](https://stackoverflow.blog/2020/06/25/how-does-spam-protection-work-on-stack-exchange/) `Stack Exchange` `2020`

8. [Auto Content Moderation in C2C e-Commerce](https://www.usenix.org/conference/opml20/presentation/ueta) `Mercari` `2020`

9. [Blocking Slack Invite Spam With Machine Learning](https://slack.engineering/blocking-slack-invite-spam-with-machine-learning/) `Slack` `2020`

10. [Cloudflare Bot Management: Machine Learning and More](https://blog.cloudflare.com/cloudflare-bot-management-machine-learning-and-more/) `Cloudflare` `2020`

11. [Anomalies in Oil Temperature Variations in a Tunnel Boring Machine](https://www.youtube.com/watch?v=YV\_uLLhPRAk) `SENER` `2020`

12. [Using Anomaly Detection to Monitor Low-Risk Bank Customers](https://www.youtube.com/watch?v=MExokMM\_Bp4&t=3s) `Rabobank` `2020`

13. [Fighting fraud with Triplet Loss](https://tech.olx.com/fighting-fraud-with-triplet-loss-86e5f79c7a3e) `OLX Group` `2020`

14. [Facebook is Now Using AI to Sort Content for Quicker Moderation](https://www.theverge.com/2020/11/13/21562596/facebook-ai-moderation) ([Alternative](https://venturebeat.com/2020/11/13/facebooks-redoubled-ai-efforts-wont-stop-the-spread-of-harmful-content/)) `Facebook` `2020`

15. How AI is getting better at detecting hate speech [Part 1](https://ai.facebook.com/blog/how-ai-is-getting-better-at-detecting-hate-speech/), [Part 2](https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/), [Part 3](https://ai.facebook.com/blog/training-ai-to-detect-hate-speech-in-the-real-world/), [Part 4](https://ai.facebook.com/blog/how-facebook-uses-super-efficient-ai-models-to-detect-hate-speech/) `Facebook` `2020`

16. [Project RADAR: Intelligent Early Fraud Detection System with Humans in the Loop](https://eng.uber.com/project-radar-intelligent-early-fraud-detection/) `Uber` `2022`

## Graph

1. [Building The LinkedIn Knowledge Graph](https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph) `LinkedIn` `2016`

2. [Scaling Knowledge Access and Retrieval at Airbnb](https://medium.com/airbnb-engineering/scaling-knowledge-access-and-retrieval-at-airbnb-665b6ba21e95) `Airbnb` `2018`

3. [Graph Convolutional Neural Networks for Web-Scale Recommender Systems](https://arxiv.org/abs/1806.01973) ([Paper](https://arxiv.org/pdf/1806.01973.pdf))`Pinterest` `2018`

4. [Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations](https://eng.uber.com/uber-eats-graph-learning/) `Uber` `2019`

5. [AliGraph: A Comprehensive Graph Neural Network Platform](https://arxiv.org/abs/1902.08730) ([Paper](https://arxiv.org/pdf/1902.08730.pdf)) `Alibaba` `2019`

6. [Contextualizing Airbnb by Building Knowledge Graph](https://medium.com/airbnb-engineering/contextualizing-airbnb-by-building-knowledge-graph-b7077e268d5a) `Airbnb` `2019`

7. [Retail Graph — Walmart’s Product Knowledge Graph](https://medium.com/walmartlabs/retail-graph-walmarts-product-knowledge-graph-6ef7357963bc) `Walmart` `2020`

8. [Traffic Prediction with Advanced Graph Neural Networks](https://deepmind.com/blog/article/traffic-prediction-with-advanced-graph-neural-networks) `DeepMind` `2020`

9. [SimClusters: Community-Based Representations for Recommendations](https://dl.acm.org/doi/10.1145/3394486.3403370) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3394486.3403370), [Video](https://crossminds.ai/video/5f3369790576dd25aef288d5/)) `Twitter` `2020`

10. [Metapaths guided Neighbors aggregated Network for Heterogeneous Graph Reasoning](https://arxiv.org/abs/2103.06474) ([Paper](https://arxiv.org/pdf/2103.06474.pdf)) `Alibaba` `2021`

11. [Graph Intention Network for Click-through Rate Prediction in Sponsored Search](https://arxiv.org/abs/2103.16164) ([Paper](https://arxiv.org/pdf/2103.16164.pdf)) `Alibaba` `2021`

12. [JEL: Applying End-to-End Neural Entity Linking in JPMorgan Chase](https://ojs.aaai.org/index.php/AAAI/article/view/17796) ([Paper](https://www.aaai.org/AAAI21Papers/IAAI-21.DingW.pdf)) `JPMorgan Chase` `2021`

13. [How AWS uses graph neural networks to meet customer needs](https://www.amazon.science/blog/how-aws-uses-graph-neural-networks-to-meet-customer-needs) `Amazon` `2022`

## Optimization

1. [Matchmaking in Lyft Line (Part 1)](https://eng.lyft.com/matchmaking-in-lyft-line-9c2635fe62c4) [(Part 2)](https://eng.lyft.com/matchmaking-in-lyft-line-691a1a32a008) [(Part 3)](https://eng.lyft.com/matchmaking-in-lyft-line-part-3-d8f9497c0e51) `Lyft` `2016`

2. [The Data and Science behind GrabShare Carpooling](https://ieeexplore.ieee.org/document/8259801) [(Part 1)](https://engineering.grab.com/the-data-and-science-behind-grabshare-part-i) (\*\*PAPER NEEDED\*\*) `Grab` `2017`

3. [How Trip Inferences and Machine Learning Optimize Delivery Times on Uber Eats](https://eng.uber.com/uber-eats-trip-optimization/) `Uber` `2018`

4. [Next-Generation Optimization for Dasher Dispatch at DoorDash](https://doordash.engineering/2020/02/28/next-generation-optimization-for-dasher-dispatch-at-doordash/) `DoorDash` `2020`

5. [Optimization of Passengers Waiting Time in Elevators Using Machine Learning](https://www.youtube.com/watch?v=vXndCC89BCw&t=4s) `Thyssen Krupp AG` `2020`

6. [Think Out of The Package: Recommending Package Types for E-commerce Shipments](https://www.amazon.science/publications/think-out-of-the-package-recommending-package-types-for-e-commerce-shipments) ([Paper](https://assets.amazon.science/0c/6c/9d0986b94bef92d148f0ac0da1ea/think-out-of-the-package-recommending-package-types-for-e-commerce-shipments.pdf)) `Amazon` `2020`

7. [Optimizing DoorDash’s Marketing Spend with Machine Learning](https://doordash.engineering/2020/07/31/optimizing-marketing-spend-with-ml/) `DoorDash` `2020`

8. [Using learning-to-rank to precisely locate where to deliver packages](https://www.amazon.science/blog/using-learning-to-rank-to-precisely-locate-where-to-deliver-packages) ([Paper](https://assets.amazon.science/69/8d/2249945a4e10ba8fc758f7523b0c/getting-your-package-to-the-right-place-supervised-machine-learning-for-geolocation.pdf))`Amazon` `2021`

## Information Extraction

1. [Unsupervised Extraction of Attributes and Their Values from Product Description](https://www.aclweb.org/anthology/I13-1190/) ([Paper](https://www.aclweb.org/anthology/I13-1190.pdf)) `Rakuten` `2013`

2. [Using Machine Learning to Index Text from Billions of Images](https://dropbox.tech/machine-learning/using-machine-learning-to-index-text-from-billions-of-images) `Dropbox` `2018`

3. [Extracting Structured Data from Templatic Documents](https://ai.googleblog.com/2020/06/extracting-structured-data-from.html) ([Paper](https://www.aclweb.org/anthology/I13-1190.pdf)) `Google` `2020`

4. [AutoKnow: self-driving knowledge collection for products of thousands of types](https://www.amazon.science/publications/autoknow-self-driving-knowledge-collection-for-products-of-thousands-of-types) ([Paper](https://arxiv.org/pdf/2006.13473.pdf), [Video](https://crossminds.ai/video/5f3369730576dd25aef288a6/)) `Amazon` `2020`

5. [One-shot Text Labeling using Attention and Belief Propagation for Information Extraction](https://arxiv.org/abs/2009.04153) ([Paper](https://arxiv.org/pdf/2009.04153.pdf)) `Alibaba` `2020`

6. [Information Extraction from Receipts with Graph Convolutional Networks](https://nanonets.com/blog/information-extraction-graph-convolutional-networks/) `Nanonets` `2021`

## Weak Supervision

1. [Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale](https://dl.acm.org/doi/abs/10.1145/3299869.3314036) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3299869.3314036)) `Google` `2019`

2. [Osprey: Weak Supervision of Imbalanced Extraction Problems without Code](https://dl.acm.org/doi/abs/10.1145/3329486.3329492) ([Paper](https://ajratner.github.io/assets/papers/Osprey\_DEEM.pdf)) `Intel` `2019`

3. [Overton: A Data System for Monitoring and Improving Machine-Learned Products](https://arxiv.org/abs/1909.05372) ([Paper](https://arxiv.org/pdf/1909.05372.pdf)) `Apple` `2019`

4. [Bootstrapping Conversational Agents with Weak Supervision](https://www.aaai.org/ojs/index.php/AAAI/article/view/5011) ([Paper](https://arxiv.org/pdf/1812.06176.pdf)) `IBM` `2019`

## Generation

1. [Better Language Models and Their Implications](https://openai.com/blog/better-language-models/) ([Paper](https://cdn.openai.com/better-language-models/language\_models\_are\_unsupervised\_multitask\_learners.pdf))`OpenAI` `2019`

2. [Image GPT](https://openai.com/blog/image-gpt/) ([Paper](https://cdn.openai.com/papers/Generative\_Pretraining\_from\_Pixels\_V2.pdf), [Code](https://github.com/openai/image-gpt)) `OpenAI` `2019`

3. [Language Models are Few-Shot Learners](https://arxiv.org/abs/2005.14165) ([Paper](https://arxiv.org/pdf/2005.14165.pdf)) ([GPT-3 Blog post](https://openai.com/blog/openai-api/)) `OpenAI` `2020`

4. [Deep Learned Super Resolution for Feature Film Production](https://graphics.pixar.com/library/SuperResolution/) ([Paper](https://graphics.pixar.com/library/SuperResolution/paper.pdf)) `Pixar` `2020`

5. [Unit Test Case Generation with Transformers](https://arxiv.org/pdf/2009.05617.pdf) `Microsoft` `2021`

## Audio

1. [Improving On-Device Speech Recognition with VoiceFilter-Lite](https://ai.googleblog.com/2020/11/improving-on-device-speech-recognition.html) ([Paper](https://arxiv.org/pdf/2009.04323.pdf))`Google` `2020`

2. [The Machine Learning Behind Hum to Search](https://ai.googleblog.com/2020/11/the-machine-learning-behind-hum-to.html) `Google` `2020`

## Validation and A/B Testing

1. [Overlapping Experiment Infrastructure: More, Better, Faster Experimentation](https://research.google/pubs/pub36500/) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36500.pdf)) `Google` `2010`

2. [The Reusable Holdout: Preserving Validity in Adaptive Data Analysis](https://ai.googleblog.com/2015/08/the-reusable-holdout-preserving.html) ([Paper](https://science.sciencemag.org/content/sci/349/6248/636.full.pdf)) `Google` `2015`

3. [Twitter Experimentation: Technical Overview](https://blog.twitter.com/engineering/en\_us/a/2015/twitter-experimentation-technical-overview.html) `Twitter` `2015`

4. [It’s All A/Bout Testing: The Netflix Experimentation Platform](https://netflixtechblog.com/its-all-a-bout-testing-the-netflix-experimentation-platform-4e1ca458c15) `Netflix` `2016`

5. [Building Pinterest’s A/B Testing Platform](https://medium.com/pinterest-engineering/building-pinterests-a-b-testing-platform-ab4934ace9f4) `Pinterest` `2016`

6. [Experimenting to Solve Cramming](https://blog.twitter.com/engineering/en\_us/topics/insights/2017/Experimenting-To-Solve-Cramming.html) `Twitter` `2017`

7. [Building an Intelligent Experimentation Platform with Uber Engineering](https://eng.uber.com/experimentation-platform/) `Uber` `2017`

8. [Scaling Airbnb’s Experimentation Platform](https://medium.com/airbnb-engineering/https-medium-com-jonathan-parks-scaling-erf-23fd17c91166) `Airbnb` `2017`

9. [Meet Wasabi, an Open Source A/B Testing Platform](https://www.intuit.com/blog/technology/engineering/meet-wasabi-an-open-source-ab-testing-platform/) ([Code](https://github.com/intuit/wasabi)) `Intuit` `2017`

10. [Analyzing Experiment Outcomes: Beyond Average Treatment Effects](https://eng.uber.com/analyzing-experiment-outcomes/) `Uber` `2018`

11. [Under the Hood of Uber’s Experimentation Platform](https://eng.uber.com/xp/) `Uber` `2018`

12. [Constrained Bayesian Optimization with Noisy Experiments](https://research.fb.com/publications/constrained-bayesian-optimization-with-noisy-experiments/) ([Paper](https://arxiv.org/pdf/1706.07094.pdf)) `Facebook` `2018`

13. [Reliable and Scalable Feature Toggles and A/B Testing SDK at Grab](https://engineering.grab.com/feature-toggles-ab-testing) `Grab` `2018`

14. [Modeling Conversion Rates and Saving Millions Using Kaplan-Meier and Gamma Distributions](https://better.engineering/modeling-conversion-rates-and-saving-millions-of-dollars-using-kaplan-meier-and-gamma-distributions/) ([Code](https://github.com/better/convoys)) `Better` `2019`

15. [Detecting Interference: An A/B Test of A/B Tests](https://engineering.linkedin.com/blog/2019/06/detecting-interference--an-a-b-test-of-a-b-tests) `LinkedIn` `2019`

16. [Announcing a New Framework for Designing Optimal Experiments with Pyro](https://eng.uber.com/oed-pyro-release/) ([Paper](https://papers.nips.cc/paper/9553-variational-bayesian-optimal-experimental-design.pdf)) ([Paper](https://arxiv.org/pdf/1911.00294.pdf)) `Uber` `2020`

17. [Enabling 10x More Experiments with Traveloka Experiment Platform](https://medium.com/traveloka-engineering/enabling-10x-more-experiments-with-traveloka-experiment-platform-8cea13e952c) `Traveloka` `2020`

18. [Large Scale Experimentation at Stitch Fix](https://multithreaded.stitchfix.com/blog/2020/07/07/large-scale-experimentation/) ([Paper](http://proceedings.mlr.press/v89/schmit19a/schmit19a.pdf)) `Stitch Fix` `2020`

19. [Multi-Armed Bandits and the Stitch Fix Experimentation Platform](https://multithreaded.stitchfix.com/blog/2020/08/05/bandits/) `Stitch Fix` `2020`

20. [Experimentation with Resource Constraints](https://multithreaded.stitchfix.com/blog/2020/11/18/virtual-warehouse/) `Stitch Fix` `2020`

21. [Computational Causal Inference at Netflix](https://netflixtechblog.com/computational-causal-inference-at-netflix-293591691c62) ([Paper](https://arxiv.org/pdf/2007.10979.pdf)) `Netflix` `2020`

22. [Key Challenges with Quasi Experiments at Netflix](https://netflixtechblog.com/key-challenges-with-quasi-experiments-at-netflix-89b4f234b852) `Netflix` `2020`

23. [Making the LinkedIn experimentation engine 20x faster](https://engineering.linkedin.com/blog/2020/making-the-linkedin-experimentation-engine-20x-faster) `LinkedIn` `2020`

24. [Our Evolution Towards T-REX: The Prehistory of Experimentation Infrastructure at LinkedIn](https://engineering.linkedin.com/blog/2020/our-evolution-towards-t-rex--the-prehistory-of-experimentation-i) `LinkedIn` `2020`

25. [How to Use Quasi-experiments and Counterfactuals to Build Great Products](https://engineering.shopify.com/blogs/engineering/using-quasi-experiments-counterfactuals) `Shopify` `2020`

26. [Improving Experimental Power through Control Using Predictions as Covariate](https://doordash.engineering/2020/06/08/improving-experimental-power-through-control-using-predictions-as-covariate-cupac/) `DoorDash` `2020`

27. [Supporting Rapid Product Iteration with an Experimentation Analysis Platform](https://doordash.engineering/2020/09/09/experimentation-analysis-platform-mvp/) `DoorDash` `2020`

28. [Improving Online Experiment Capacity by 4X with Parallelization and Increased Sensitivity](https://doordash.engineering/2020/10/07/improving-experiment-capacity-by-4x/) `DoorDash` `2020`

29. [Leveraging Causal Modeling to Get More Value from Flat Experiment Results](https://doordash.engineering/2020/09/18/causal-modeling-to-get-more-value-from-flat-experiment-results/) `DoorDash` `2020`

30. [Iterating Real-time Assignment Algorithms Through Experimentation](https://doordash.engineering/2020/12/08/optimizing-real-time-algorithms-experimentation/) `DoorDash` `2020`

31. [Spotify’s New Experimentation Platform (Part 1)](https://engineering.atspotify.com/2020/10/29/spotifys-new-experimentation-platform-part-1/) [(Part 2)](https://engineering.atspotify.com/2020/11/02/spotifys-new-experimentation-platform-part-2/) `Spotify` `2020`

32. [Interpreting A/B Test Results: False Positives and Statistical Significance](https://netflixtechblog.com/interpreting-a-b-test-results-false-positives-and-statistical-significance-c1522d0db27a) `Netflix` `2021`

33. [Interpreting A/B Test Results: False Negatives and Power](https://netflixtechblog.com/interpreting-a-b-test-results-false-negatives-and-power-6943995cf3a8) `Netflix` `2021`

34. [Running Experiments with Google Adwords for Campaign Optimization](https://doordash.engineering/2021/02/05/google-adwords-campaign-optimization/) `DoorDash` `2021`

35. [The 4 Principles DoorDash Used to Increase Its Logistics Experiment Capacity by 1000%](https://doordash.engineering/2021/09/21/the-4-principles-doordash-used-to-increase-its-logistics-experiment-capacity-by-1000/) `DoorDash` `2021`

36. [Experimentation Platform at Zalando: Part 1 - Evolution](https://engineering.zalando.com/posts/2021/01/experimentation-platform-part1.html) `Zalando` `2021`

37. [Designing Experimentation Guardrails](https://medium.com/airbnb-engineering/designing-experimentation-guardrails-ed6a976ec669) `Airbnb` `2021`

38. [How Airbnb Measures Future Value to Standardize Tradeoffs](https://medium.com/airbnb-engineering/how-airbnb-measures-future-value-to-standardize-tradeoffs-3aa99a941ba5) `Airbnb` `2021`

38. [Network Experimentation at Scale](https://research.fb.com/publications/network-experimentation-at-scale/)([Paper](https://arxiv.org/abs/2012.08591)] `Facebook` `2021`

39. [Universal Holdout Groups at Disney Streaming](https://medium.com/disney-streaming/universal-holdout-groups-at-disney-streaming-2043360def4f) `Disney` `2021`

40. [Experimentation is a major focus of Data Science across Netflix](https://netflixtechblog.com/experimentation-is-a-major-focus-of-data-science-across-netflix-f67923f8e985) `Netflix` `2022`

41. [Search Journey Towards Better Experimentation Practices](https://engineering.atspotify.com/2022/02/search-journey-towards-better-experimentation-practices/) `Spotify` `2022`

42. [Artificial Counterfactual Estimation: Machine Learning-Based Causal Inference at Airbnb](https://medium.com/airbnb-engineering/artificial-counterfactual-estimation-ace-machine-learning-based-causal-inference-at-airbnb-ee32ee4d0512) `Airbnb` `2022`

43. [Challenges in Experimentation](https://eng.lyft.com/challenges-in-experimentation-be9ab98a7ef4) `Lyft` `2022`

44. [Overtracking and Trigger Analysis: Reducing sample sizes while INCREASING sensitivity](https://booking.ai/overtracking-and-trigger-analysis-how-to-reduce-sample-sizes-and-increase-the-sensitivity-of-71755bad0e5f) `Booking` `2022`

45. [Meet Dash-AB — The Statistics Engine of Experimentation at DoorDash](https://doordash.engineering/2022/05/24/meet-dash-ab-the-statistics-engine-of-experimentation-at-doordash/) `DoorDash` `2022`

46. [Comparing quantiles at scale in online A/B-testing](https://engineering.atspotify.com/2022/03/comparing-quantiles-at-scale-in-online-a-b-testing) `Spotify` `2022`

## Model Management

1. [Operationalizing Machine Learning—Managing Provenance from Raw Data to Predictions](https://vimeo.com/274396495) `Comcast` `2018`

2. [Overton: A Data System for Monitoring and Improving Machine-Learned Products](https://arxiv.org/abs/1909.05372) ([Paper](https://arxiv.org/pdf/1909.05372.pdf)) `Apple` `2019`

3. [Runway - Model Lifecycle Management at Netflix](https://www.usenix.org/conference/opml20/presentation/cepoi) `Netflix` `2020`

4. [Managing ML Models @ Scale - Intuit’s ML Platform](https://www.usenix.org/conference/opml20/presentation/wenzel) `Intuit` `2020`

5. [ML Model Monitoring - 9 Tips From the Trenches](https://building.nubank.com.br/ml-model-monitoring-9-tips-from-the-trenches/) `Nubank` `2021`

## Efficiency

1. [GrokNet: Unified Computer Vision Model Trunk and Embeddings For Commerce](https://ai.facebook.com/research/publications/groknet-unified-computer-vision-model-trunk-and-embeddings-for-commerce/) ([Paper](https://scontent-sea1-1.xx.fbcdn.net/v/t39.8562-6/99353320\_565175057533429\_3886205100842024960\_n.pdf?\_nc\_cat=110&\_nc\_sid=ae5e01&\_nc\_ohc=WQBaZy1gnmUAX8Ecqtt&\_nc\_ht=scontent-sea1-1.xx&oh=cab2f11dd9154d817149cb73e8b692a8&oe=5F5A3778)) `Facebook` `2020`

2. [How We Scaled Bert To Serve 1+ Billion Daily Requests on CPUs](https://blog.roblox.com/2020/05/scaled-bert-serve-1-billion-daily-requests-cpus/) `Roblox` `2020`

3. [Permute, Quantize, and Fine-tune: Efficient Compression of Neural Networks](https://arxiv.org/abs/2010.15703) ([Paper](https://arxiv.org/pdf/2010.15703.pdf)) `Uber` `2021`

## Ethics

1. [Building Inclusive Products Through A/B Testing](https://engineering.linkedin.com/blog/2020/building-inclusive-products-through-a-b-testing) ([Paper](https://arxiv.org/pdf/2002.05819.pdf)) `LinkedIn` `2020`

2. [LiFT: A Scalable Framework for Measuring Fairness in ML Applications](https://engineering.linkedin.com/blog/2020/lift-addressing-bias-in-large-scale-ai-applications) ([Paper](https://arxiv.org/pdf/2008.07433.pdf)) `LinkedIn` `2020`

3. [Introducing Twitter’s first algorithmic bias bounty challenge](https://blog.twitter.com/engineering/en\_us/topics/insights/2021/algorithmic-bias-bounty-challenge) `Twitter` `2021`

3. [Examining algorithmic amplification of political content on Twitter](https://blog.twitter.com/en\_us/topics/company/2021/rml-politicalcontent) `Twitter` `2021`

3. [A closer look at how LinkedIn integrates fairness into its AI products](https://engineering.linkedin.com/blog/2022/a-closer-look-at-how-linkedin-integrates-fairness-into-its-ai-pr) `LinkedIn` `2022`

## Infra

1. [Reengineering Facebook AI’s Deep Learning Platforms for Interoperability](https://ai.facebook.com/blog/reengineering-facebook-ais-deep-learning-platforms-for-interoperability) `Facebook` `2020`

2. [Elastic Distributed Training with XGBoost on Ray](https://eng.uber.com/elastic-xgboost-ray/) `Uber` `2021`

## MLOps Platforms

1. [Meet Michelangelo: Uber’s Machine Learning Platform](https://eng.uber.com/michelangelo-machine-learning-platform/) `Uber` `2017`

2. [Operationalizing Machine Learning—Managing Provenance from Raw Data to Predictions](https://vimeo.com/274396495) `Comcast` `2018`

3. [Big Data Machine Learning Platform at Pinterest](https://www.slideshare.net/Alluxio/pinterest-big-data-machine-learning-platform-at-pinterest) `Pinterest` `2019`

4. [Core Modeling at Instagram](https://instagram-engineering.com/core-modeling-at-instagram-a51e0158aa48) `Instagram` `2019`

5. [Open-Sourcing Metaflow - a Human-Centric Framework for Data Science](https://netflixtechblog.com/open-sourcing-metaflow-a-human-centric-framework-for-data-science-fa72e04a5d9) `Netflix` `2019`

6. [Managing ML Models @ Scale - Intuit’s ML Platform](https://www.usenix.org/conference/opml20/presentation/wenzel) `Intuit` `2020`

7. [Real-time Machine Learning Inference Platform at Zomato](https://www.youtube.com/watch?v=0-3ES1vzW14) `Zomato` `2020`

8. [Introducing Flyte: Cloud Native Machine Learning and Data Processing Platform](https://eng.lyft.com/introducing-flyte-cloud-native-machine-learning-and-data-processing-platform-fb2bb3046a59) `Lyft` `2020`

9. [Building Flexible Ensemble ML Models with a Computational Graph](https://doordash.engineering/2021/01/26/computational-graph-machine-learning-ensemble-model-support/) `DoorDash` `2021`

10. [LyftLearn: ML Model Training Infrastructure built on Kubernetes](https://eng.lyft.com/lyftlearn-ml-model-training-infrastructure-built-on-kubernetes-aef8218842bb) `Lyft` `2021`

11. ["You Don't Need a Bigger Boat": A Full Data Pipeline Built with Open-Source Tools](https://github.com/jacopotagliabue/you-dont-need-a-bigger-boat) ([Paper](https://arxiv.org/abs/2107.07346)) `Coveo` `2021`

12. [MLOps at GreenSteam: Shipping Machine Learning](https://neptune.ai/blog/mlops-at-greensteam-shipping-machine-learning-case-study) `GreenSteam` `2021`

13. [Evolving Reddit’s ML Model Deployment and Serving Architecture](https://www.reddit.com/r/RedditEng/comments/q14tsw/evolving\_reddits\_ml\_model\_deployment\_and\_serving/) `Reddit` `2021`

14. [Redesigning Etsy’s Machine Learning Platform](https://www.etsy.com/codeascraft/redesigning-etsys-machine-learning-platform/) `Etsy` `2021`

15. [Building a Platform for Serving Recommendations at Etsy](https://www.etsy.com/codeascraft/building-a-platform-for-serving-recommendations-at-etsy) `Etsy` `2022`

16. [Intelligent Automation Platform: Empowering Conversational AI and Beyond at Airbnb](https://medium.com/airbnb-engineering/intelligent-automation-platform-empowering-conversational-ai-and-beyond-at-airbnb-869c44833ff2) `Airbnb` `2022`

17. [DARWIN: Data Science and Artificial Intelligence Workbench at LinkedIn](https://engineering.linkedin.com/blog/2022/darwin--data-science-and-artificial-intelligence-workbench-at-li) `LinkedIn` `2022`

18. [The Magic of Merlin: Shopify's New Machine Learning Platform](https://shopify.engineering/merlin-shopify-machine-learning-platform) `Shopify` `2022`

19. [Zalando's Machine Learning Platform](https://engineering.zalando.com/posts/2022/04/zalando-machine-learning-platform.html) `Zalando` `2022`

20. [Inside Meta's AI optimization platform for engineers across the company](https://ai.facebook.com/blog/looper-meta-ai-optimization-platform-for-engineers/) ([Paper](https://arxiv.org/pdf/2110.07554.pdf)) `Meta` `2022`

21. [Monzo’s machine learning stack](https://monzo.com/blog/2022/04/26/monzos-machine-learning-stack) `Monzo` `2022`

22. [Evolution of ML Fact Store](https://netflixtechblog.com/evolution-of-ml-fact-store-5941d3231762) `Netflix` `2022`

23. [Deployment for Free: A Machine Learning Platform for Stitch Fix's Data Scientists](https://multithreaded.stitchfix.com/blog/2022/07/14/deployment-for-free/) `Stitch Fix`

## Practices

1. [Practical Recommendations for Gradient-Based Training of Deep Architectures](https://arxiv.org/abs/1206.5533) ([Paper](https://arxiv.org/pdf/1206.5533.pdf)) `Yoshua Bengio` `2012`

2. [Machine Learning: The High Interest Credit Card of Technical Debt](https://research.google/pubs/pub43146/) ([Paper](https://storage.googleapis.com/pub-tools-public-publication-data/pdf/43146.pdf)) ([Paper](https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf)) `Google` `2014`

3. [Rules of Machine Learning: Best Practices for ML Engineering](https://developers.google.com/machine-learning/guides/rules-of-ml) `Google` `2018`

4. [On Challenges in Machine Learning Model Management](http://sites.computer.org/debull/A18dec/p5.pdf) `Amazon` `2018`

5. [Machine Learning in Production: The Booking.com Approach](https://booking.ai/https-booking-ai-machine-learning-production-3ee8fe943c70) `Booking` `2019`

6. [150 Successful Machine Learning Models: 6 Lessons Learned at Booking.com](https://booking.ai/150-successful-machine-learning-models-6-lessons-learned-at-booking-com-681e09107bec) ([Paper](https://dl.acm.org/doi/pdf/10.1145/3292500.3330744)) `Booking` `2019`

7. [Successes and Challenges in Adopting Machine Learning at Scale at a Global Bank](https://www.youtube.com/watch?v=QYQKG5OcwEI) `Rabobank` `2019`

8. [Challenges in Deploying Machine Learning: a Survey of Case Studies](https://arxiv.org/abs/2011.09926) ([Paper](https://arxiv.org/pdf/2011.09926.pdf)) `Cambridge` `2020`

9. [Reengineering Facebook AI’s Deep Learning Platforms for Interoperability](https://ai.facebook.com/blog/reengineering-facebook-ais-deep-learning-platforms-for-interoperability) `Facebook` `2020`

10. [The problem with AI developer tools for enterprises](https://towardsdatascience.com/the-problem-with-ai-developer-tools-for-enterprises-and-what-ikea-has-to-do-with-it-b26277841661) `Databricks` `2020`

11. [Continuous Integration and Deployment for Machine Learning Online Serving and Models](https://eng.uber.com/continuous-integration-deployment-ml/) `Uber` `2021`

12. [Tuning Model Performance](https://eng.uber.com/tuning-model-performance/) `Uber` `2021`

13. [Maintaining Machine Learning Model Accuracy Through Monitoring](https://doordash.engineering/2021/05/20/monitor-machine-learning-model-drift/) `DoorDash` `2021`

14. [Building Scalable and Performant Marketing ML Systems at Wayfair](https://www.aboutwayfair.com/careers/tech-blog/building-scalable-and-performant-marketing-ml-systems-at-wayfair) `Wayfair` `2021`

15. [Our approach to building transparent and explainable AI systems](https://engineering.linkedin.com/blog/2021/transparent-and-explainable-AI-systems) `LinkedIn` `2021`

16. [5 Steps for Building Machine Learning Models for Business](https://shopify.engineering/building-business-machine-learning-models) `Shopify` `2021`

17. [Data Is An Art, Not Just A Science—And Storytelling Is The Key](https://shopifyengineering.myshopify.com/blogs/engineering/data-storytelling-shopify) `Shopify` `2022`

18. [Best Practices for Real-time Machine Learning: Alerting](https://building.nubank.com.br/best-practices-for-real-time-machine-learning-alerting/) `Nubank` `2022`

19. [Automatic Retraining for Machine Learning Models: Tips and Lessons Learned](https://building.nubank.com.br/automatic-retraining-for-machine-learning-models/) `Nubank` `2022`

## Team structure

1. [Engineers Shouldn’t Write ETL: A Guide to Building a High Functioning Data Science Department](https://multithreaded.stitchfix.com/blog/2016/03/16/engineers-shouldnt-write-etl/) `Stitch Fix` `2016`

2. [Building The Analytics Team At Wish](https://medium.com/wish-engineering/scaling-analytics-at-wish-619eacb97d16) `Wish` `2018`

3. [Beware the Data Science Pin Factory: The Power of the Full-Stack Data Science Generalist](https://multithreaded.stitchfix.com/blog/2019/03/11/FullStackDS-Generalists/) `Stitch Fix` `2019`

4. [Cultivating Algorithms: How We Grow Data Science at Stitch Fix](https://cultivating-algos.stitchfix.com) `Stitch Fix`

5. [Analytics at Netflix: Who We Are and What We Do](https://netflixtechblog.com/analytics-at-netflix-who-we-are-and-what-we-do-7d9c08fe6965) `Netflix` `2020`

6. [Building a Data Team at a Mid-stage Startup: A Short Story](https://erikbern.com/2021/07/07/the-data-team-a-short-story.html) `Erikbern` `2021`

7. [A Behind-the-Scenes Look at How Postman’s Data Team Works](https://entrepreneurshandbook.co/a-behind-the-scenes-look-at-how-postmans-data-team-works-fded0b8bfc64) `Postman` `2021`

8. [Data Scientist x Machine Learning Engineer Roles: How are they different? How are they alike?](https://building.nubank.com.br/data-scientist-x-machine-learning-engineer-roles-how-are-they-different-how-are-they-alike/) `Nubank` `2022`

## Fails

1. [When It Comes to Gorillas, Google Photos Remains Blind](https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/) `Google` `2018`

2. [160k+ High School Students Will Graduate Only If a Model Allows Them to](http://positivelysemidefinite.com/2020/06/160k-students.html) `International Baccalaureate` `2020`

3. [An Algorithm That ‘Predicts’ Criminality Based on a Face Sparks a Furor](https://www.wired.com/story/algorithm-predicts-criminality-based-face-sparks-furor/) `Harrisburg University` `2020`

4. [It's Hard to Generate Neural Text From GPT-3 About Muslims](https://twitter.com/abidlabs/status/1291165311329341440) `OpenAI` `2020`

5. [A British AI Tool to Predict Violent Crime Is Too Flawed to Use](https://www.wired.co.uk/article/police-violence-prediction-ndas) `United Kingdom` `2020`

6. More in [awful-ai](https://github.com/daviddao/awful-ai)

<br>

\*\*P.S., Want a summary of ML advancements?\*\* Get up to speed with survey papers 👉[`ml-surveys`](https://github.com/eugeneyan/ml-surveys)