

Andrew O. Arnold

andrew.arnold@gmail.com ♦ andrewoarnold.com ♦ [Google Scholar](https://scholar.google.com/citations?user=...) ♦ [LinkedIn](https://www.linkedin.com/in/andrewoarnold/)

PROFILE

A hands-on technical and scientific expert with experience both as a tech lead and senior individual contributor delivering complex, high impact applied machine learning research projects in the areas of science, technology and quantitative trading. Currently Principal Applied Machine Learning Engineer at Shopify, building machine learning based features and products to help make commerce better for everyone. Former Chief Scientist at Oracle Alpha, leading machine learning and natural language understanding research for systematic fundamental hedge fund. Previously, at AWS, built and led a team of 45+ applied research scientists, working closely with engineering and product partners, to identify, solve, build and launch large scale, cross-org machine learning, natural language understanding and time series problems for our customers. Before that, held various leadership positions in the fields of machine learning, artificial intelligence, and quantitative trading, at firms such as Google, Cubist/Point72 and WorldQuant. Adjunct faculty in Financial Engineering at New York University.

EDUCATION

Carnegie Mellon University, *School of Computer Science, Machine Learning Department*, Advisor: **William W. Cohen**

Ph.D. Machine Learning: *Exploiting Domain and Task Regularities for Robust Named Entity Recognition*, 2009

M.Sc. Knowledge Discovery & Data Mining: *A Comparison of Methods for Transductive Transfer Learning*, 2007

Developed unsupervised transfer learning techniques for natural language processing and structure comprehension.

Columbia University, *Columbia College*

B.A. Computer Science/Artificial Intelligence, Advisor: **Salvatore J. Stolfo**, 2003

Magna Cum Laude, Phi Beta Kappa, John Jay Scholar, Computer Science Departmental Award, Dean's List

Developed one-class support vector machine techniques for non-signature intrusion detection.

EXPERIENCE

Shopify

June 2024 – Present

Principal Applied Machine Learning Engineer

New York, NY, USA

Building machine learning based features and products to help make commerce better for everyone.

New York University

Fall 2018 – Present

Adjunct Professor

New York, NY, USA

Teaching courses related to machine learning, natural language processing and quantitative trading.

Oracle Alpha

July 2022 – April 2024

Chief Scientist

New York, NY, USA

Led machine learning and natural language processing research and production for emerging systematic fundamental hedge fund.

Amazon Web Services (AWS)

March 2020 – July 2022

Senior Science Manager, AI Labs

New York, NY, USA

Built and led a team of 45+ applied research scientists, working closely with our product and engineering partners, to build novel AI-enabled cloud services driven by large scale machine learning and natural language understanding research, specifically in the areas of Large Language Models and No-Code/Low-Code Application Development. Developed and led projects from start to launch, with complex requirements and requiring technical leadership and coordination across multiple teams and organizations.

- [Amazon Q Developer \(formerly CodeWhisperer\)](#): ML-powered coding companion (similar to [GitHub Copilot](#))
- [Amazon Kendra](#): Enterprise search powered by machine learning
- [Amazon Contact Lens](#): Contact center call sentiment and analytics powered by machine learning
- [Amazon QuickSight Q](#): Natural language business intelligence querying
- [Amazon Lookout for Metrics](#): Automatic time series anomaly detection and diagnosis

Google

July 2018 – March 2020

Research

New York, NY, USA

Developed novel, large scale machine learning (ML) and natural language processing (NLP) algorithms with a focus on robustness and transfer learning, advancing the state-of-the-art and creating significant business impact in multiple, billion dollar products.

Cubist Systematic Strategies (Point72)

June 2015 – June 2018

Portfolio Manager; Research Director: Advanced Methods Group

New York, NY, USA

Created first of its kind hybrid research and trading group focused on researching, developing and trading proprietary machine learning based signals and strategies, with particular focus on bridging the gap from theoretical results to trading performance.

- Developing high sharpe, high capacity, high and medium frequency futures signals and strategies, managing all aspects of the research, development and trading processes, including idea generation, validation, optimization and execution.
- Creating novel machine learning techniques as well as adapting existing academic techniques to the trading domain.
- Building top-tier machine learning research team. Developing machine learning recruitment pipeline from scratch. Recruiting, hiring, training, managing and developing collaborative team of tier-1 machine learning researchers. Establishing

effective recruiting presence at top tier conferences and universities. Encouraging team development and close academic relationships by hosting firm-wide seminars and collaborations with high profile academic speakers.

- Providing machine learning leadership and expertise throughout the firm, helping to introduce other PM's to novel techniques along with guidance on how to use those techniques most effectively in a trading environment.

Ophir Partners, LLC Co-founder; Partner; Portfolio Manager
July 2013 – February 2015 New York, NY, USA

Built and co-ran all research and trading aspects of high sharpe, high capacity medium frequency equities quantitative hedge fund. Developed core alphas, strategies and platform based on diverse set of data sources, signal construction and portfolio optimization techniques. Co-managed research team. Focused on robust, differentiated, diverse, neutral, and scalable sources of alpha and risk.

TrexQuant Management, LLC Partner; Portfolio Manager; Chief Technology Officer
Fall 2011 – May 2013 Stamford, CT, USA

Co-built and co-ran all research and trading aspects of \$1B high sharpe, high capacity medium frequency equities quantitative hedge fund, including alphas, strategies and platform. Designed, built and managed all research and technology. Co-managed team of researchers in collaborative environment.

WorldQuant, LLC (Millennium Partners) Portfolio Manager; Alpha Researcher
June 2009 – August 2011 New York, NY, USA

Researched, developed and traded profitable medium frequency quantitative equity portfolios. Developed techniques for combining disparate trading signals into highly uncorrelated high-sharpe portfolios. Extracted novel features from large, noisy structured and unstructured datasets and developed into uncorrelated, high-sharpe trading signals. Researched and sourced novel data sources.

Merrill Lynch (QSA) Summer Associate, High Frequency Proprietary Trading (QSA/GSRG, former ETL)
Summer 2008 New York, NY, USA

Researched, developed and tested novel natural language and statistical techniques for modeling news feeds into tradeable signals on a high-frequency quantitative equities prop trading desk. Investigated performance benefits of non-linear learning methods.

Microsoft Research Asia Research Intern, Web Search and Mining Group
Summer 2007 Beijing, China; Supervisor: **Hang Li**

Developed semi-supervised and transfer learning based methods for improving internet search through query-dependent ranking.

IBM Research Research Intern, Data Analytics & Math, Thomas J. Watson Center
Summer 2006 Yorktown Heights, NY, USA; Supervisors: **Chid Apte, Naoki Abe**

Developed techniques for learning temporal causal networks for anomaly detection, corporate profiling, and process engineering.

Google Software Engineer (full time), Local Search/Maps
March 2004 – July 2004 New York, NY, USA; Supervisor: **Craig Nevill-Manning**

Developed automated location submission and verification system for increased coverage and accuracy of early local.google.com

Bloomberg L.P. Researcher & Developer, Network Design & Information Security
June 2003 – February 2004 New York, NY, USA

Designed and developed security technologies and policies to increase performance of 50,000+ access-points network.

Microsoft Corporation Intern, Natural Language Group
Summer 2002 Redmond, WA, USA; Supervisor: **Maria Katsova**

Researched and developed nlp transformation based learning algorithm to improve relevancy and decrease cost of search.

Lehman Brothers Japan Assistant Database Programmer, Database Group
Summer 2001 Tokyo, Japan

Developed database and web information systems used in critical business operations throughout Asia.

PUBLICATIONS

Jing Wang, Jie Shen, Xiaofei Ma and **Andrew O. Arnold**. "Uncertainty-based Active Learning for Reading Comprehension." *Transactions on Machine Learning Research (TMLR)*, 2022.

Yuantong Li, Xiaokai Wei, Zijian Wang, Shen Wang, Parminder Bhatia, Xiaofei Ma and **Andrew O. Arnold**. "Debiasing Neural Retrieval via In-batch Balancing Regularization." *NAACL Workshop on Gender Bias in Natural Language Processing (NAACL:GeBNLP)*, 2022.

Zhihan Zhou, Dejiao Zhang, Wei Xiao, Nicholas Dingwall, Xiaofei Ma, **Andrew O. Arnold** and Bing Xiang. "Learning Dialogue Representations from Consecutive Utterances." *North American Chapter of Association for Computational Linguistics (NAACL)*, 2022.

Xisen Jin, Dejiao Zhang, Henghui Zhu, Wei Xiao, Shang-Wen Li, Xiaokai Wei, **Andrew O. Arnold** and Xiang Ren. "**Lifelong Pretraining: Continually Adapting Language Models to Emerging Corpora.**" *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022.

Danilo Neves Ribeiro, Shen Wang, Xiaofei Ma, Xiaokai Wei, Henghui Zhu, Rui Dong, Xinchu Chen, Peng Xu, Zhiheng Huang, **Andrew O. Arnold** and Dan Roth. "**Entailment Tree Explanations via Iterative Retrieval-Generation Reasoner.**" *Findings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2022.

Zheng Li, Zijian Wang, Ming Tan, Ramesh Nallapati, Parminder Bhatia, **Andrew O. Arnold**, Dan Roth and Bing Xiang. "**DQ-BART: Efficient Sequence-to-Sequence Model via Joint Distillation and Quantization.**" *Association for Computational Linguistics (ACL)*, 2022.

Mufan Sang, Haoqi Li, Fang Liu, **Andrew O. Arnold** and Li Wan. "**Self-Supervised Speaker Verification with Simple Siamese Network and Self-Supervised Regularization.**" *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022.

Shen Wang, Xiaokai Wei, Cicero Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, **Andrew O. Arnold** and Philip S. Yu. "**Knowledge Graph Representation via Hierarchical Hyperbolic Neural Graph Embedding.**" *IEEE International Conference on Big Data (BigData)*, 2021.

Dejiao Zhang, Wei Xiao, Henghui Zhu, Xiaofei Ma and **Andrew O. Arnold**. "**Virtual Augmentation Supported Contrastive Learning of Sentence Representations.**" *Findings of the Association for Computational Linguistics (ACL)*, 2022.

Andy T. Liu, Wei Xiao, Henghui Zhu, Dejiao Zhang, Shang-Wen Li and **Andrew O. Arnold**. "**QaNER: Prompting Question Answering Models for Few-shot Named Entity Recognition.**" *arXiv:2203.01543*, 2022.

Xiaokai Wei, Shen Wang, Dejiao Zhang, Parminder Bhatia and **Andrew O. Arnold**. "**Knowledge Enhanced Pretrained Language Models: A Comprehensive Survey.**" *arXiv:2110.08455*, 2021.

Dejiao Zhang, Shang-Wen Li, Wei Xiao, Henghui Zhu, Ramesh Nallapati, **Andrew O. Arnold** and Bing Xiang. "**Pairwise Supervised Contrastive Learning of Sentence Representations.**" *Empirical Methods in Natural Language Processing (EMNLP)*, 2021.

Yifan Gao, Henghui Zhu, Patrick Ng, Cicero Nogueira dos Santos, Zhiguo Wang, Feng Nan, Dejiao Zhang, Ramesh Nallapati, **Andrew O. Arnold** and Bing Xiang. "**Answering Ambiguous Questions through Generative Evidence Fusion and Round-Trip Prediction.**" *Association for Computational Linguistics (ACL)*, 2021.

Feng Nan, Cicero Nogueira dos Santos, Henghui Zhu, Patrick Ng, Kathleen McKeown, Ramesh Nallapati, Dejiao Zhang, Zhiguo Wang, **Andrew O. Arnold** and Bing Xiang. "**Improving Factual Consistency of Abstractive Summarization via Question Answering.**" *Association for Computational Linguistics (ACL)*, 2021.

Xiaofei Ma, Cicero Nogueira dos Santos and **Andrew O. Arnold**. "**Contrastive Fine-tuning Improves Robustness for Neural Rankers.**" *Findings of the Association for Computational Linguistics (ACL)*, 2021.

Dejiao Zhang, Feng Nan, Xiaokai Wei, Shang-Wen Li, Henghui Zhu, Kathleen McKeown, Ramesh Nallapati, **Andrew O. Arnold** and Bing Xiang. "**Supporting Clustering with Contrastive Learning.**" *North American Chapter of the Association for Computational Linguistics (NAACL)*, 2021.

Shen Wang, Xiaokai Wei, Cicero Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, **Andrew O. Arnold**, Bing Xiang, Isabel F. Cruz and Philip S. Yu. "**Mixed-Curvature Multi-relational Graph Neural Network for Knowledge Graph Completion.**" *The Web Conference (WWW)*, 2021.

Andrew O. Arnold and William W. Cohen. "**Instance-based Transfer Learning for Multilingual Deep Retrieval.**" *The Web Conference Workshop on Multilingual Search (WWW)*, 2021.

Haitian Sun, **Andrew O. Arnold**, Tania Bedrax-Weiss, Fernando Pereira and William W. Cohen. "**Faithful Embeddings for Knowledge Base Queries.**" *Neural Information Processing Systems (NeurIPS)*, 2020.

Cheng Tang and **Andrew O. Arnold**. "**Neural document expansion for ad-hoc information retrieval.**" *arXiv:2012.14005*, 2020.

Andrew O. Arnold and William W. Cohen. "**Information Extraction as Link Prediction: Using Curated Citation Networks to Improve Gene Detection.**" *AAAI Conference on Weblogs and Social Media (ICWSM)*, 2009.

Amr Ahmed, **Andrew O. Arnold**, Luis Pedro Coelho, Joshua Kangas, Abdul-Saboor Sheikh, Eric Xing, William Cohen and Robert F. Murphy. "**Structured Literature Image Finder.**" *ISMB BioLINK Special Interest Group (BioLINK)*, 2009.

Andrew O. Arnold and William W. Cohen. "**Intra-document Structural Frequency Features for Semi-supervised Domain Adaptation.**" *Conference on Information and Knowledge Management (CIKM)*, 2008.

Xiubo Geng, Tie-Yan Liu, Tao Qin, **Andrew O. Arnold**, Hang Li and Harry Shum. "**Query Dependent Ranking Using K-Nearest Neighbor.**" *International Conference on Information Retrieval (SIGIR)*, 2008. U.S. patent application 20100169323 A1.

Andrew O. Arnold, Ramesh Nallapati and William W. Cohen. "**Exploiting Feature Hierarchy for Transfer Learning in Named Entity Recognition.**" *Association for Computational Linguistics: Human Language Technologies (ACL:HLT)*, 2008.

Andrew O. Arnold, Ramesh Nallapati and William W. Cohen. “**A Comparative Study of Methods for Transductive Transfer Learning.**” *International Conference on Data Mining Workshop on Mining and Management of Biological Data (ICDM)*, 2007.

Andrew O. Arnold, Yan Liu and Naoki Abe. “**Temporal Causal Modeling with Graphical Granger Methods.**” *International Conference on Knowledge Discovery and Data Mining (KDD)*, 2007.

Andrew O. Arnold, Joseph E. Beck and Richard Scheines. “**Feature Discovery in the Context of Educational Data Mining: An Inductive Approach.**” *AAAI Workshop on Educational Data Mining (AAAI)*, 2006.

Kristinn R. Thórisson, Hrvoje Benko, Denis Abramov, **Andrew O. Arnold**, Sameer Maskey, and Aruchunan Vaseekaran. “**Constructionist Design Methodology for Interactive Intelligences.**” In *AI Magazine, (AAAI)*, 2004.

Eleazar Eskin, **Andrew O. Arnold**, Michael Prerau, Leonid Portnoy and Sal Stolfo. “**A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data.**” In Daniel Barbara and Sushil Jajodia (editors), *Applications of Data Mining in Computer Security*, Kluwer, 2002. Incorporated into U.S. patent 8544087 B1.

PATENTS

Two undisclosed Amazon patents, pending review, 2020 – 2022.

Constrained prefix matching for generating next token predictions. US patent application #20230418567 A1, 2023. Work done at Amazon Web Services.

Programmatically generating evaluation data sets for code generation models. US patent application #20230418566 A1, 2023. Work done at Amazon Web Services.

Random token segmentation for training next token prediction models. US patent application #20230419036 A1, 2023. Work done at Amazon Web Services.

Validating and providing proactively generated code suggestions. US patent application #20230418565 A1, 2023. Work done at Amazon Web Services.

Methods of unsupervised anomaly detection using a geometric framework. U.S. patent granted #8544087 B1, 2013. Work done at Columbia University.

Query-Dependent Ranking Using K-Nearest Neighbor. US patent application #20100169323 A1, 2008. Work done at Microsoft Research Asia.

INVITED TALKS

"A Shallow Introduction to Deep AI for Finance." Wolfe Research, AI Disruption in Investment Management Conference. New York, NY (May 22, 2023).

"ChatGPT, NLP, Predictive Analytics: Is Artificial Intelligence Finally Here?" Battle of the Quants, Panel with Li Deng (Vatic/Citadel). New York, NY (May 10, 2023).

"Recruiting for the Next Generation of Quant." Carnegie Mellon University's Master of Science in Computational Finance (MSCF) 25th Anniversary Celebration. Pittsburgh, PA (March 21, 2020) [Postponed].

"Building the Optimal Data/ML Teams." AI & Data Science in Trading. New York, NY (March 17-18, 2020) [Postponed].

"Transfer Learning for Machine Learning and NLP: Adapting Models to Changing Markets." AI & Data Science in Trading. New York, NY (March 17-18, 2020) [Postponed].

"Navigating Data Challenges with the Latest ML/AI Applications." AI & Data Science in Trading. New York, NY (March 17-18, 2020) [Postponed].

"AI in the Workplace," CMU NYC Tech & Entrepreneurship Panel, Liquidnet, New York, NY (September 26, 2019).

"Machine Learning Developments and Applications to Quantitative Trading," AI & Data Science in Trading. New York, NY (March 19-20, 2019) [did not attend].

"Transfer Learning for Quantitative Trading," machineByte 2018: The Global Machine Learning in Quantitative Investment Management Forum. Half Moon Bay, CA (December 13, 2018) [did not attend].

"Machine Learning and Trading," Career Speaker Series. Bendheim Center for Finance, Princeton University, Princeton, NJ (March 29, 2017).

"Exploiting Document Structure and Feature Hierarchy for Semi-supervised Domain Adaptation." Machine Learning Lunch. Carnegie Mellon University, Pittsburgh, PA (September 29, 2008).

"A Comparison of Methods for Transductive Transfer Learning." Information Retrieval and Mining Seminar. Microsoft Research Asia, Beijing, China (May 30, 2007).

"Feature Discovery in the Context of Educational Data Mining: An Inductive Approach." IBM Mathematical Sciences Department Seminar. IBM Watson Research, Yorktown Heights, NY (July 6, 2006).

"Causal Modeling for Anomaly Detection." IBM Mathematical Sciences Department 2006 Summer Student Seminar Series. IBM Watson Research, Yorktown Heights, NY (June 23, 2006).

COURSES TAUGHT

- *News Analytics and Machine Learning Strategies for Quantitative Trading*, New York University (2018 – Present).
- *Machine Learning* with **Tom Mitchell** and **William W. Cohen**, Carnegie Mellon University.
- *Multimedia Databases and Data Mining* with **Christos Faloutsos**, Carnegie Mellon University.
- *Honors Introduction to Computer Science* with **John R. Kender**; *Distinguished Teaching Assistant* Award, Columbia University.

SKILLS

Computer Languages: Python, PyTorch, TensorFlow, Matlab, Java, C/C++/#, R, Scala, Perl, SQL, AWS, GCP, Airflow
Registrations: Series 7/63 (lapsed)

AWARDS & HONORS

2009	<i>Finalist, Elsevier Grand Challenge: Knowledge Enhancement in the Life Sciences</i>
2008	<i>Presidential Fellow in the Life Sciences, Richard King Mellon Foundation</i>
2003	<i>Computer Science Department Award, Columbia University</i>
2002	<i>Distinguished Teaching Assistant Award, Columbia University, Computer Science Department</i>
1999 – 2003	<i>John Jay Scholar, Columbia University</i>
1999 – 2003	<i>Dean's List, Columbia College, Columbia University</i>

SERVICE

Volunteer machine learning consultant	Solutions Journalism Network
Area Chair	North American Chapter Association Computational Linguistics (NAACL)
Program Committee	Conference on Neural Information Processing Systems (NeurIPS) International Conference On Machine Learning (ICML) World Wide Web Conference (WWW) European Conference on Machine Learning (ECML) Principles and Practice of Knowledge Discovery in Databases (PKDD)
Member	American Association for Artificial Intelligence (AAAI) Association for Computational Linguistics (ACL) Association for Computing Machinery (ACM)

HOBBIES & INTERESTS

- (Automated) poker
- Vinyl records