

EDO LIBERTY

CONTACT	<i>email:</i> edo.liberty@gmail.com <i>web:</i> www.edoliberty.com	
EXPERIENCE	Pinecone <i>Founder, CEO</i> Pinecone is an AI SaaS company offering Vector Database services. Amazon AI, AWS <i>Director of Research, Head of Amazon AI Labs</i> Lead a science-engineering organization building machine learning algorithms into SageMaker, Elastic Search, Kinesis, Glue, QuickSight, Personalize, and Forecast. Amazon AI, AWS <i>Senior Manager of Research for Amazon SageMaker</i> Hired and managed Amazon AI Algorithms Group, launched Amazon SageMaker Elastic Algorithms. Yahoo Research, New York <i>Head of Yahoo Research in New York</i> Managed research efforts for Yahoo Finance, Yahoo Sports, Yahoo Mail, and Yahoo's advertising Platform. Yahoo Labs, New York <i>Director of Research</i> Hired and managed Yahoo's Scalable Machine Learning group. Built scalable machine learning solutions for Yahoo critical applications. Yahoo Labs, Israel <i>Manager of Research for Yahoo Mail</i> Driving machine learning based functionality and monetization. Cognitive Media Networks <i>Co-Founder</i> Automatic content recognition for connected televisions Obtained 80% of the US market, acquired by VISIO	<i>May 2019 - Today</i> <i>Mar 2018 - Apr 2019</i> <i>Aug 2016 - Mar 2018</i> <i>Mar 2016 - July 2016</i> <i>Aug 2013 - Mar 2016</i> <i>Nov 2009 - Aug 2013</i> <i>May 2008 - Nov 2009</i>
EDUCATION	Yale University <i>Post Doctoral Fellow</i> Member of the program in Applied Mathematics Yale University <i>Ph.D. in Computer Science</i> Thesis title: Accelerated Dense Random Projection Advised by Steven W. Zucker, funded by NGA, AFOSR and NSF. Google Research, New York <i>Research Intern</i> University of California, UCLA <i>NSF Research fellow</i> Google, Israel <i>Software Engineering Intern</i> Tel Aviv University <i>B.Sc.</i> Physics and Computer Science, double major (magna cum laude)	<i>Jan 2009 - Oct 2009</i> <i>Sep 2004 - May 2009</i> <i>Jun 2008 - Aug 2008</i> <i>Sep 2007 - Dec 2007</i> <i>Jun 2007 - Aug 2007</i> <i>Oct 2000 - Jun 2003</i>
TEACHING	Tel Aviv University <i>Lecturer, Algorithms in Data Mining</i> Streaming and online algorithms in data mining Tel Aviv University <i>Lecturer, Algorithms in Data Mining</i> Randomized algorithms in data mining of massive data sets Tel Aviv University <i>Lecturer, Data Mining</i> Applied algorithms in data mining of massive data sets Yale University <i>Teaching assistant, Data Mining</i> Hands-on data mining basics and exercises	<i>Oct 2012 - Feb 2013</i> <i>Oct 2011 - Feb 2012</i> <i>Oct 2010 - Feb 2011</i> <i>Oct 2006 - Feb 2008</i>
SERVICE	Program Committee, Senior Program Committee or Area Chair for: ICML 2019, KDD 2019, COLT 2019, KDD 2018, APPROX 2018, SODA 2018, ESA 2017, KDD 2017, WWW 2017, WebMail 2017, WSDM 2017, AISTATS 2017, SODA 2016, ICDM 2016, ICML 2016 (Outstanding Reviewer Award), KDD 2016, COLT, 2015, IC2S2 2015, WSDM 2015, BIG DATA 2014, WWW 2014 SODA 2014, WSDM 2013, ESA 2011	
AWARDS	Best Papers at: PODS 2021, KDD 2013, TechPulse 2012, and SODA 2011. More than \$500k in academic grants, scholarships, and fellowships.	

INVITED KEYNOTES AND TUTORIALS	The Pinecone Vector Database System
	Andy Pavlo's CMU Database Seminar, 2021
	Apache DataSketches and mining data streams
	San Fransisco Data Council, July 2020
	Benefits and Challenges of combining Deep Learning and Retrieval Tasks
	Future of Information and Communication, 2020
	Streaming Algorithms
	ICML, Time Series Workshop, June 2019
	Mergeable Summaries and the DataSketches Library
	Southern Data Science Conference, Atlanta 2019
	Machine Learning Systems
	with Alex Smola; KDD 2018
	Mergeable Summaries and the DataSketches Library
	Network Traffic Measurement and Analysis Conference, TMA 2018
	Infinitely Scalable Machine Learning in Amazon SageMaker
CONFERENCE PUBLICATIONS	Network Traffic Measurement and Analysis Conference, TMA Experts Summit 2018
	Correlation Clustering: From Theory to Practice
	with Francesco Bonchi, and David García-Soriano; KDD 2014
	Machine Learning at Yahoo
	Yahoo TechPulse tutorial 2013
	Data Mining in Streams
	with Jelani Nelson; KDD 2012
	Data Mining in the Streaming Model: Approximating Massive Matrices
	IBM Machine Learning Day 2012
	Fast Random Projections, theory and practice
	14th Annual Workshop on Applied and Computational Mathematics
	Fast Random Projections survey and new results
	Institute of Advanced Studies (IAS) at Princeton
	Projective Clustering Product Quantization
	Aditya Krishnan, Edo Liberty: Arxiv 2021
	Streaming Quantiles Algorithms with Small Space and Update Time
	Nikita Ivkin, Edo Liberty, Kevin Lang, Zohar Karnin, Vladimir Braverman: Sensors 2022
	Relative Error Streaming Quantiles
	Graham Cormode, Zohar Karnin, Edo Liberty, Justin Thaler, Pavel Veselý: Best Paper at PODS 2021
	From the lab to production: A case study of session-based recommendations in the home-improvement domain
	Pigi Kouki, Ilias Fountalis, Nikolaos Vasiloglou, Xiquan Cui, Edo Liberty, Khalifeh Al Jadda: RecSys 2020
	Amazon SageMaker Elastic Algorithms
	Edo Liberty, Zohar Karnin, Bing Xiang, Laurence Rouesnel, Baris Coskun, Ramesh Nallapati, Julio Delgado Mangas, Amir Sadoughi, Yury Astashonok, Piali Das, Can Balioglu, Saswata Chakravarty, Madhav Jha, Philip Gautier, Tim Januschowski, Valentin Flunkert, David Arpin, and Alex Smola: SIGMOD 2020
	Coresets, Discrepancy, and Sketches in Machine Learning
	Zohar Karnin and Edo Liberty: COLT 2019
	Asymmetric Random Projections
	Nicholas Ryder, Zohar Karnin, and Edo Liberty: Arxiv 2019
	Proxquant: Quantized neural networks via proximal operators
	Yu Bai, Yu-Xiang Wang, Edo Liberty: ICLR 2019
	Optimal Message Scheduling for Aggregation
	Leyuan Wang, Mu Li, Edo Liberty, Alex Smola: MLSYS 2018

A High-Performance Algorithm for Identifying Frequent Items in Data Streams
Daniel Anderson, Pryce Bevan, Kevin Lang, Edo Liberty, L. Rhodes, Justin Thaler: IMC 2017

Greedy Minimization of Weakly Supermodular Set Functions
Edo Liberty, Maxim Sviridenko: APPROX 2017

Optimal Quantile Approximation in Streams
Zohar Karnin, Kevin Lang, Edo Liberty; FOCS 2016

An Efficient Frequent Directions Algorithm for Sparse Matrices
Mina Ghashami, Edo Liberty, Jeff M. Phillips; KDD 2016

Stratified Sampling meets Machine Learning
Kevin Lang, Edo Liberty, Konstantin Shmakov; ICML 2016

Space Lower Bounds for Itemset Frequency Sketches
Edo Liberty, Michael Mitzenmacher, Justin Thaler, Jonathan Ullman; PODS 2016

Online PCA with Spectral Bounds
Zohar Karnin, Edo Liberty; COLT 2015

An Algorithm for Online K-Means Clustering
Edo Liberty, Ram Sriharsha, Maxim Sviridenko; ALENEX 2015

Online Principal Component Analysis
Christos Boutsidis, Dan Garber, Zohar Karnin, Edo Liberty; SODA 2015

Correlation clustering: from theory to practice
Francesco Bonchi, David García-Soriano, Edo Liberty; KDD 2014

Near-Optimal Entrywise Sampling for Data Matrices
Dimitris Achlioptas, Zohar Karnin, Edo Liberty; NIPS 2013

Simple and Deterministic Matrix Sketches
Edo Liberty Best Paper, KDD 2013

Threading Machine Generated Email
Nir Ailon, Zohar Karnin, Edo Liberty, Yoelle Maarek; TechPulse best paper, WSDM 2013

Unsupervised SVMs: On the complexity of the Furthest Hyperplane Problem
Edo Liberty, Shachar Lovett, Omri Weinstein; COLT 2012

Framework and Algorithms for Network Bucket Testing
Liran Katzir, Edo Liberty, and Oren Somekh; WWW 2012

Almost Optimal Unrestricted Fast Johnson-Lindenstrauss Transform
Nir Ailon, Edo Liberty; Best Paper at SODA 2011

Improved Approximation Algorithms for Bipartite Correlation Clustering
Nir Ailon, Noa Avigdor-Elgrabli, Edo Liberty, Anke van Zuylen; ESA 2011

Automatically tagging email by leveraging other users' folders
Yehuda Koren, Edo Liberty, Yoelle Maarek, Roman Sandler; KDD 2011

Estimating sizes of social networks via biased sampling
Liran Katzir, Edo Liberty, Oren Somekh; WWW 2011

Inverted index compression via online document routing
Gal Lavee, Ronny Lempel, Edo Liberty; WWW 2011

Similarity kernels via bi-clustering
Minh Tam Le, John Sweeney, Edo Liberty, Steven W. Zucker; ISI 2010

Correlation Clustering Revisited: The ‘True’ Cost of Error Minimization Problems
Nir Ailon, Edo Liberty; ICALP 2009

Dense Fast Random Projections and Lean Walsh Transforms
Edo Liberty, Nir Ailon, Amit Singer; RANDOM 2008

Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes
Nir Ailon, Edo Liberty; invited to TALG, SODA 2008 Best Papers shortlist

Estimating Sizes of Social Networks via Biased Sampling

Oren Somekh; Liran Katzir, Edo Liberty, Ioana Cosma, Internet Mathematics 2013

An Almost Optimal Unrestricted Fast Johnson-Lindenstrauss Transform

Nir Ailon, Edo Liberty; TALG 2013

Improved Approximation Algorithms for Bipartite Correlation Clustering

Nir Ailon, Noa Avigdor-Elgrabli, Edo Liberty, Anke van Zuylen; to appear SICOMP

Unsupervised SVMs: On the Complexity of the Furthest Hyperplane Problem

Zohar Karnin, Edo Liberty, Shachar Lovett, Roy Schwartz, Omri Weinstein; JMLR 2012

Dense Fast Random Projections and Lean Walsh Transforms

Edo Liberty, Steven Zucker; DCG 2010

The Mailman algorithm: a note on matrix vector multiplication

Edo Liberty, Nir Ailon, Amit Singer; IPL 2009

Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes

Nir Ailon, Edo Liberty ; DCG 2008

A fast randomized algorithm for the approximation of matrices

Edo Liberty, Franco Woolfe, Vladimir Rokhlin, and Mark Tygert; ACHA 2008

Randomized algorithms for the low-rank approximation of matrices

Edo Liberty, Franco Woolfe, Gunnar Martinsson, Vladimir Rokhlin, Mark Tygert; PNAS 2007

ELECTRONIC MAIL
RELATED PATENTS

Classifying man versus machine generated email

Zohar Karnin, Guy Halawi, David Wajc, Edo Liberty

A System for Email sequence identification

Edo Liberty, Zohar Karnin, Yoelle Maarek, Natalie Aizenberg

Sponsored Apps Marketplace in eMail

Ronny Lempel, Yoelle Maarek, Edward Bortnikov, Edo Liberty

Mining Global Email Folders For Identifying Auto-folders tags

Vishwanath Ramarao, Andrei Broder, Idan Szpektor, Edo Liberty, Yehuda Koren, Mark Risher, and Yoelle Maarek

Email sequence identification

Edo Liberty ,Zohar Karnin, Yoelle Maarek

Mailing List Identification and Representation

Zohar Karnin, Michal Aharon, Edo Liberty, Yoelle Maarek

Identification of subject line templates

Zohar Karnin, Edo Liberty, David Wajc, Guy Halawi

Computerized system and method for modifying a message to apply security features to the message's content

Edo Liberty, Yoelle Maarek

Electronic message composition support method and apparatus

J Tetreault, A Pappu, E Liberty, L Cao, M Liu, E Pavlick, G Tsur, Y Maarek

Mail Lint: Write Better Emails

Joel Tetreault, Aasish Pappu, Edo Liberty

Liangliang Cao, Meizhu Liu ,Ellie Tobochnik, Gilad Tzur, Yoelle Maarek

ALGORITHMS AND
SYSTEMS PATENTS

Method And System For Clustering Data Points

Nir Ailon, Edo Liberty, Hari Khalsa

Methods for filtering data and filling in missing data using nonlinear inference

Edo Liberty, Steven Zucker, Yosi Keller, Mauro M. Maggioni, Ronald R. Coifman, Frank Geshwind, and in collaboration with Plain Sight Systems.

Generalized Stratified Sampling

Kevin Lang, Edo Liberty ,Konstantin Shmakov

On-line content sampling

KJ Lang, E Liberty, K Shmakov

System and Method for Experimentation and Deployment of Machine Learning Models on Cloud Based Platforms

Edo Liberty, Stefano Stefani, Alexander Smola, Craig Wiley, Steve Loeppky, Tom Faulhaber, Swami Sivasubramanian, Zohar Karnin

Method for post-training Hyperparameter Tuning by training Machine Learning States

Edo Liberty, Zohar Karnin

Autoscaling of Training Machine Learning Jobs on Cloud Infrastructures

Edo Liberty, Stefano Stefani, Swami Sivasubramanian, Zohar Karnin, Tom Faulhaber, Alexander Smola, Craig Wiley, Amir Sadoughi, Dayanand Rangegowda

A system for autoscaling and hosting of ML Models for production inference

Edo Liberty, Stefano Stefani, Steve Loeppky, Craig Wiley, Tom Faulhaber

Online training with delayed feedback with applications to bandwidth-efficient communication over networks

Edo Liberty, Madhav Jha

System Architecture for Container Based Large Scale Machine Learning Platforms

Stefano Stefani, Craig Wiley, Thomas Faulhaber, Alexander Smola, Steven Loeppky, Richard Bice, Edo Liberty, Swaminathan Sivasubramanian, Charles Swan, Taylor Goodhart

Method and Systems for Optimal Graph Synchronization for Distribute Machine Learning

Mu Li, Edo Liberty, Alexander Smola, Leyuan Wang

OTHER PATENTS

Training machine learning models for physical agents and robotic controls with simulations

S Genc, E Liberty - US Patent 10,824,913, 2020

Machine Learning system to remove accent from spoken speech

Edo Liberty, Leo Dirac

Methods for Displaying Contextually Targeted Content on a Connected Television

Zeev Neumeier, Edo Liberty

Methods for Identifying Video Segments and Displaying Contextually Targeted Content on Connected Televisions

Zeev Neumeier, Edo Liberty

Contest Generation Methods for Daily Fantasy Sports

Justin Thaler, Maxim Sviridenko, Edo Liberty, Prerit Uppal, Ron Belmarch, Jerry Shen

Fantasy Sports Data Analysis for Game Structure Development

Justin Thaler, Maxim Sviridenko, Edo Liberty, Prerit Uppal, Ron Belmarch, Jerry Shen

Machine Learning model-assisted real-time enhancement of audio/video over a network call to significantly lower bandwidth requirements

Madhav Jha, Edo Liberty