

# Avneesh Singh Saluja

---

## CONTACT INFORMATION

1149 Amador Ave.  
Berkeley, CA  
94707, USA

☎: +1-650-526-8792  
✉: [asaluja@gmail.com](mailto:asaluja@gmail.com)  
🌐: <http://www.cs.cmu.edu/~avneesh>

## SYNOPSIS

Applied research scientist with a background in machine learning for natural language processing (NLP). I have built systems from prototype to production, served as a technical lead for small pods of research engineers and scientists, and founded applied research teams from scratch.

## PROFESSIONAL EXPERIENCE

- 9/2018–present    **Netflix**  
Los Angeles & Los Gatos, CA  
Research Scientist, Content Machine Learning  
Staff research lead on foundational projects involving representation learning of entities based on natural language and other discrete data. Conducting novel research on fine-grained content understanding via text and video, with primary applications in content performance & valuation prediction, and content programming.
- 10/2015–9/2018    **Airbnb**  
San Francisco, CA  
Research Scientist, AI Lab  
Senior tech lead for applied research projects across multiple company initiatives, with a focus on multimodal learning for user-generated content. Worked on various projects across search, relevance, growth, infrastructure, trust & safety, and customer service.
- 6/2014–8/2014    **eBay Inc.**  
San Jose, CA  
Research Intern, Statistical Machine Translation Group  
Project: modeling compositional and non-compositional semantics, with applications in machine translation
- 5/2013–8/2013    **Microsoft Research**  
Redmond, WA  
Research Intern, Statistical Machine Translation Group  
Project: graph-based semi-supervised learning for phrase table expansion
- 5/2012–8/2012    **IBM Research T.J. Watson Center**  
Yorktown Heights, NY  
Research Intern, Statistical NLP Group  
Project: graph-based unsupervised word similarities from multiple feature types
- 7/2009–5/2010    **July Systems**  
Bangalore, India  
Lead Manager, Product & Product Marketing  
Product managed won *Best Mobile International* category at 2010 Mobile Excellence Awards
- 7/2007–3/2009    **Goldman Sachs**  
New York, NY & Hong Kong  
Financial Analyst, Structured Equity Solutions Group

## EDUCATION

### **Carnegie Mellon University**, Pittsburgh, PA

Ph.D. in Electrical & Computer Engineering, September 2015

- Dissertation: “Low-dimensional Context-dependent Translation Models”
- Advisors: Chris Dyer & Ian Lane
- Affiliated with Language Technologies Institute, School of Computer Science
- Recipient, Bertucci Fellowship (College of Engineering), 2015
- Recipient, eBay Graduate Fellowship, 2014

### **Carnegie Mellon University**, Pittsburgh, PA

M.S. in Electrical & Computer Engineering, May 2013

- Graduate coursework in statistics, machine learning, learning theory, speech recognition, probabilistic graphical models, structured prediction, optimization

### **Stanford University**, Stanford, CA

B.S. *with Distinction* in Electrical Engineering, April 2007

- Signal processing concentration
- President’s Award for Academic Excellence, Tau Beta Pi

## ADVISORY ROLES

- 11/2018–present    **Altovita**, London, U.K.  
AI Advisor  
Providing continuous guidance around the company’s core AI product
- 3/2017–present    **Sentio**, San Francisco, CA  
NLP Advisor  
Guiding the Data Science team on best practices and cutting edge approaches for their summarization, transcription, and search efforts

## PUBLICATIONS

### **Journals**

**A. Saluja** and Y. Zhang; Online Discriminative Learning for Machine Translation with Binary-valued Feedback. *Machine Translation*, Vol. 28 (2) pp. 69-90, 2014.

### **Conferences**

S. Mehta, B. Azarnoush, B. Chen, **A. Saluja**, V. Misra, B. Bahani, and R. Kumar; Simplify-then-Translate: Automatic Preprocessing for Black-Box Translation. In *Proceedings of the Association for the Advancement of Artificial Intelligence*, 2020.

C. Mitcheltree, V. Wharton, and **A. Saluja**; Using Aspect Extraction Approaches to Generate Review Summaries and User Profiles. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL) - Industry Track*, 2018.

**A. Saluja**, C. Dyer, and S. Cohen; Latent Variable Synchronous CFGs for Hierarchical Translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

A.P. Parikh, **A. Saluja**, C. Dyer, and E.P. Xing; Language Modeling with Power Low Rank Ensembles. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014. (**Best Paper Nominee**)

**A. Saluja**, H. Hassan, K. Toutanova, and C. Quirk; Graph-based Semi-Supervised Learning of Translation Models from Monolingual Data. In *Proceedings of the Association for Computational Linguistics (ACL)*, 2014.

**A. Saluja**, I. Lane, and Y. Zhang; Machine Translation with Binary Feedback: a Large-Margin Approach. In *Conference of the Association for Machine Translation in the Americas (AMTA)*, 2012.

**A. Saluja**, P. Sundararajan, and O.J. Mengshoel; Age-Layered Expectation Maximization for Parameter Learning in Bayesian Networks. In *Artificial Intelligence & Statistics (AISTATS)*, 2012.

**A. Saluja**, I. Lane, and Y. Zhang; Context-aware Language Modeling for Conversational Speech Translation. In *Machine Translation Summit XIII*, 2011.

### Workshops

G. Bhat, **A. Saluja**, M. Dye, and J. Florjanczyk; Hierarchical Encoders for Modeling and Interpreting Screenplays. In *3rd Workshop on Narrative Understanding*, 2021.

**A. Saluja** and J. Navrátil; Graph-based Unsupervised Learning of Word Similarities Using Heterogeneous Feature Types. In *TextGraphs 8: Graph-based Methods for Natural Language Processing*, 2013.

**A. Saluja** and B. Kveton; Semi-Supervised Learning with Cover Trees. In *Big Learning: 2011 NIPS Workshop on Parallel and Large-Scale Machine Learning*, 2011.

**A. Saluja**, F. Mokaya, M. Phielipp, and B. Kveton; Automatic Identity Inference for Smart Televisions. In *AAAI 2011 Workshop on Lifelong Learning*, 2011.

### Preprints & Technical Reports

**A. Saluja**, C. Dyer, and J.D. Ruvini; Paraphrase-Supervised Models of Compositionality. *arXiv:1801.10293*. February 2018 (originally February 2015).

**A. Saluja**, M. Pakdaman, D. Piao, and A.P. Parikh; Infinite Mixed Membership Matrix Factorization. *arXiv:1401.3413*. January 2014.

### INVITED TALKS

June 2016	OpenAir Conference, San Francisco: <i>Machine Learning in a Community-driven Marketplace</i>
October 2016	Facebook Inc., Menlo Park: <i>Neural Networks and NLP at Airbnb</i>
November 2016	ReWork Machine Intelligence Summit, New York: <i>Extracting Customer Insights at Airbnb</i>
November 2017	Open Data Science Conference, San Francisco: <i>Deep Learning and Language Processing at Airbnb</i>
April 2018	AI NextCon Silicon Valley, Santa Clara: <i>Deep Learning and Language Processing at Airbnb</i>
August 2019	1 <sup>st</sup> AllenNLP Summit, Seattle: <i>AllenNLP at Netflix</i>
December 2020	NeurIPS Expo: <i>NLP at Netflix</i>

PROFESSIONAL SERVICE	Co-organizer	★ Modern Machine Learning and Natural Language Processing Workshop held at NIPS 2014
	Reviewer	ACL, NAACL, EACL, EMNLP, JMLR, IWSLT, AAAI
TEACHING EXPERIENCE	8/2014–12/2014	<b>18-751</b> Applied Stochastic Processes Teaching Assistant and Guest Lecturer
	1/2013–5/2013	<b>18-799M</b> Advanced Machine Learning Teaching Assistant and Guest Lecturer
LANGUAGES	Programming:	Python, SQL, Bash
	Human:	English, Hindi, Urdu, Nepalese, Spanish (fluent); Punjabi, Bengali (conversational); Arabic (basic)
PERSONAL	Lived in 9 countries on 6 continents, with K-12 schooling in international schools MENSA International member Interests: violin, guitar, soccer, cricket, and squash	
REFERENCES	Available upon request	