

**JUDE KHOUJA** +1 202 733 0000 mkhooja@gmail.com

## SUMMARY

**NLP:** Language representation, LM, QA, Multi-task learning, Computational Social Science

**ML:** Linear Models, SVM, Random Forests, PCA, Clustering, LDA. **Programing:** Python, Pytorch, SQL

**Other:** Mentoring ML Scientists and Research Interns, Technical Lead, Recruiting

## EDUCATION

**Stanford University** - Graduate coursework in Natural Language Processing with Deep Learning (CS224n), 2017

**University of Maryland College Park** - Masters in Information Management, May 2012

**Damascus University** - B.E. Computer Science, Focus: Artificial Intelligence, July 2007

**HONORS** Fulbright Scholarship, U.S. Department of State 2010.

## PUBLICATIONS

Khouja, J. Stance Prediction and Claim Verification: An Arabic Perspective. Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER) workshop at ACL 2020

Social Media for Political Change: The Activists, Governments and Firms Triangle of Powers During the Arab Movement. Handbook of Research on Political Activism in the Information Age (pages 26-36). IGI Global 2015

Knowledge and Theme Discovery across Very Large Biological Data Sets Using Distributed Queries: A Prototype Combining Unstructured and Structured Data. PLOS 2013

Mr. LDA: A Flexible Large Scale Topic Modeling Package using Variational Inference in MapReduce. ACM International Conference on World Wide Web, 2012.

## HIGHLIGHTED EXPERIENCE

### SR. STAFF APPLIED SCIENTIST, FORETHOUGHT.AI

Feb 2021 - Present

Working on conversational modeling, intent detection and question answering. Managing partnerships for data annotation and research collaboration. Hiring and mentoring a talented team of engineers and scientists in NLP and ML.

### SR. PRINCIPAL ML SCIENTIST, SAGE INTACCT

Apr 2020 - Jan 2021

Helped the company release their first AI product by revamping all ML models (Improve relative F1-score 60-80%). Helped build the ML engineering team and hired key positions.

### SENIOR APPLIED SCIENTIST, MICROSOFT

May 2018 - Sep 2019

Contributed to scaling up Language Model distributed training to tens of billions of words and explored the use of sub-word representations (PBE, Wordpiece) in speech services.

### SENIOR /PRINCIPAL DATA SCIENTIST, SALESFORCE

Sep 2015 - Feb 2018

Drove the team's NLP DL foundation and automated business processes by building text classification models for the customer support processes. Developed reusable machine learning pipeline and feature engineering libraries.

### DATA SCIENCE LEAD, CRITTERCISM

Jul 2014 - Nov 2014

Applied clustering techniques and text matching for cardinality reduction.

Built internet facing analytics portal for tracking live mobile performance metrics worldwide.

Built AWS on demand analytics infrastructure using S3 and EMR.

### DATA ANALYTICS LEAD, IREX

Nov 2013 - Jun 2014

Built and managed data driven software, team and processes for documenting human rights violations.

Oversaw the design and implementation of video and image based annotation system.

### BIG DATA SPECIALIST, ORACLE

Mar 2013 - Oct 2013

Developed machine learning proof of concepts in the public sector using technologies including Hadoop, Hive, Pig, R Enterprise, Mahout and other proprietary and open source tools.

### DATA SCIENTIST, ORACLE

Jun 2012 - Aug 2012

Was the Lead Data Scientist in the Oracle/NCI partnership project which resulted in winning the "2012 Best Government Big Data Solution" Award.

Developed MapReduce programs in Java and Python for analyzing simulated Gene data of 900 million patients.

Assisted in building a 9-node Hadoop (CDH) cluster.

### GRADUATE RESEARCH ASSISTANT, UMIACS

Oct 2011 - May 2012

Evaluated large scalable distributed Topic Modeling algorithm (LDA) and applied them for unsupervised lexicon expansion.

Developed an Arabic version of the Word Count tool for sentiment analysis and honor dictionary validation.

### PROGRAM OFFICER - M&E, UNITED NATIONS

Aug 2009 - Aug 2010

Developed and implemented a comprehensive M&E plan for a \$6 Million grant by The Global Fund Program (GFATM).

Oversaw the design and implementation of the health management information system.

### TECHNOLOGY CONSULTANT, UNITED NATIONS

Aug 2008 - Mar 2009

Conducted comprehensive assessment of the existing IT infrastructure and capacity for the United Nations (UNDP) project on Aid Effectiveness and produced a full analysis report to the Minister of state planning commission.

## CERTIFICATION

M&E for UN programs, United Nations, Geneva 2009.

SCJP (Sun Certified Java Programmer), Sun 2008.