# Alex Morales | CV

(805) 705 1364 • aermslo@outlook.com • amorale4.github.io

My interests lie in the intersection of Machine Learning, Natural Language Processing and Text Mining. In the past I have focused on **Text Mining for Social Media Analysis**, **Community Question Answering**, **Identifying Quality Content in Online Communities** and **Misinformation Detection**. I have collaborated on interdisciplinary projects that touch societal problems related to health. I am excited about work in areas such as Information Integrity, Abuse Detection, Truth Discovery and User Behavior Modeling.

## Work Experience

**Twitter - Misleading Information Detection Team**
Machine Learning Engineer *July 2021 - Present*
Developed and deployed machine learning models for misleading information detection. Formalized and implemented claim features which led to a 33% increase in precision for misleading information detection. Created machine learning model measurement pipelines for the team's detection models for both offline and online settings. Contributed to the detection architecture to enable A/B experimentation for misinformation detection.

**Adobe Research - Systems Technology Lab**
Data Scientist Intern *Summer 2014*
Mentor: Dr. Nedim Lipka
Temporal Summarization on Real-Time Social Media Streams. Developed an event summarization prototype system on Twitter using Spark and GraphX.

## Education

**University of Illinois at Urbana-Champaign**
PhD Computer Science, August 2014 - December 2020, Advisor: Prof. ChengXiang Zhai
Thesis Topic: *Model-based Feature Construction and Text Representation for Social Media Analysis*
Committee Members: Jiawei Han, Julia Hockenmaier, Lyle Ungar

**University of California, Santa Barbara**
M.S. Computer Science, June 2014, Advisor: Prof. Xifeng Yan
Thesis Topic: *Detecting Synthetic Review Spam*

**University of California, Santa Barbara**
B.S. Computer Science, June 2012

## Research Experience

**Categorizing Severity in COVID-19 Misinformation** **DAIS @ UIUC**
TIMan Group *Spring 2020 - Fall 2020*
Categorized the granularity of severity in COVID-19 misinformation using social media posts.
Developed misinformation data collection and annotation scheme, and created a website to streamline this process.
Helped develop several deep-learning model baselines using BERT, HAN, and deFEND.

**Identifying Inspector Bias via Inspector Report Consistency Analysis** **DAIS @ UIUC**
TIMan Group, in collaboration with the Champaign-Urbana Public Health Department *Spring 2020 - Fall 2020*
Worked with the CU-PHD to identify inconsistencies and biases in health inspection reports by applying text mining methods.
Developed a text classification method for re-inspection prediction solely based on inspector comments.
Performed inspector experience bias analysis based on temporal scoring patterns over an archive of reports from 2008-2018.

**Reliable Question Answering in Online Discussion Forums** **DAIS @ UIUC**
TIMan Group *Spring 2018 - Spring 2019*
Lead the creation of a novel community discussion forum dataset by crawling several ask communities on Reddit.com.
Proposed and developed a user-aspect reliability model using both user topical expertise and comment context information in order to identify trustworthy and reliable comments from users with diverse expertise and perform user-expertise identification.

**HIV Radar Project** **Social Action Lab @ UIUC**
Interdisciplinary research, PI: Prof. Dolores Albarracin *Fall 2015 - Fall 2020*

Analyzed social media influences on HIV/STI transmission behaviors by combining Twitter messages and socio-demographic variables for risk factor discovery. Helped develop the Twitter crawlers, maintained the MongoDB server to store the tweets, and developed several related predictive risk factor models using social media data.

**Automatic Humor Identification in User Reviews** **DAIS @ UIUC**
TIMan Group *Fall 2014 - Spring 2016*
Proposed and developed model to capture incongruity, unexpectedness, and language differential semantics in reviews from the Yelp Dataset Challenge to identify humorous intent and high quality reviews.

## Teaching Experience

**MOOC - Text Mining and Analysis/Data Mining Capstone**
Teaching Assistant @UIUC *Summer 2015, Summer 2019*
Instructor: Prof. ChengXiang Zhai
Helped develop the first offering of the Data Mining Capstone Course offered online for the MCS-DS degree. Created the online quizzes and exams used by Coursera. Developed the data mining capstone project and grading rubrics.

**CS290D - Advanced Data Mining**
Teaching Assistant @UCSB *Spring 2014*
Instructor: Prof. Xifeng Yan
Sole TA for the first offering of the graduate Data Mining course. Gave guest lectures, and graded students' assignments and exams.

**CS165A - Artificial Intelligence**
Teaching Assistant @UCSB *Spring 2013*
Instructor: Prof. Xifeng Yan
Lead the TAs to ensure the large undergraduate course was managed efficiently.

## Awards

Hack Week **#HackForGood** Award for developing tooling to improve consistent content moderation, Twitter 2022
**FDO-Managed Retail Program Standards Grant**, UIUC 2019-2020
**Graduate College Distinguished Fellowship**, UIUC 2015-2018
**Support for Under-Represented Groups in Engineering PhD Fellowship**, UIUC 2014-2017
**GEM PhD Fellowship, Adobe Fellow**, The National GEM Consortium, UIUC 2014
**Hispanic College Fund Scholarship, Ford Scholar**, UCSB 2011
**Hispanic College Fund Scholarship, Google Scholar** , UCSB 2010
**College of Engineering Scholar** - UCSB 2009

## Other Activities/Service

**CIKM Applied Research Program Committee Member**, CIKM, Summer 2020, Summer 2021, Summer 2022
**Summer Research Experience for Undergraduates Mentor**, CS REU @ UIUC, Summer 2020
**Illinois Scholars Undergraduate Research Mentor**, ISUR @ UIUC, Fall 2017 - Spring 2018
**Sloan University Center of Exemplary Mentoring at Illinois, Peer Mentor**, UCEM @ UIUC, Fall 2016 - Spring 2017
**Mentoring Undergraduates in Science & Engineering**, MUSE @ UIUC, Fall 2016 - Spring 2017
**CS Graduate Ambassador**, CS @ UIUC, Spring 2015
**Multicultural Engineering Recruitment for Graduate Education, Graduate Ambassador**, MERGE @ UIUC, Spring 2014, 2015, 2016, 2017, 2018

## Technical Skills

**Programming Languages, Libraries and Tools:** Python, Java, Scala, Scalding, Hadoop, Aurora, Matlab, C/C++, SQL/mySQL, UNIX Shell Scripting, LaTeX, HTML, MongoDB, Spark, pandas, Dataflow, Airflow, Vertex AI, BigQuery, BQML, GCP, Looker Studio, Jupyter Notebook
**Deep Learning and Other packages:** pyTorch, TensorFlow, Torchtext, Keras, scikit-learn

## Publications


[1] Arkin Dharawat, Ismini Lourentzou, **Alex Morales**, and ChengXiang Zhai. Drink bleach or do what now? covid-hera: A study of risk-informed health decision making in the presence of covid-19 misinformation. *Proceedings of the International AAAI Conference on Web and Social Media*, 16(1):1218–1227, May 2022.

[2] Tre Tomaszewski, **Alex Morales**, Ismini Lourentzou, Rachel Caskey, Bing Liu, Alan Schwartz, Jessie Chin, et al. Identifying false human papillomavirus (hpv) vaccine information and corresponding risk perceptions from twitter: advanced predictive models. *Journal of medical Internet research*, 23(9):e30451, 2021.

[3] Man-pui Sally Chan, **Alex Morales**, Maria Zlotorzynska, Patrick Sullivan, Travis Sanchez, Chengxiang Zhai, and Dolores Albarracín. Estimating the influence of twitter on pre-exposure prophylaxis use and hiv testing as a function of rates of men who have sex with men in the united states. *AIDS*, 35:S101–S109, 2021.

[4] **Alex Morales**. *Model-based Feature Construction and Text Representation for Social Media Analysis*. PhD thesis, University of Illinois at Urbana-Champaign, 2020.

[5] **Alex Morales**, Kanika Narang, Hari Sundaram, and ChengXiang Zhai. Crowdqm: Learning aspect-level user reliability and comment trustworthiness in discussion forums. In *The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2020.

[6] Nupoor Gandhi, **Alex Morales**, Sally Man-Pui Chan, Dolores Albarracin, and ChengXiang Zhai. Predicting opioid overdose crude rates with text-based twitter features (student abstract). In *the AAAI Conference on Artificial Intelligence. Vol. 33*, 2020.

[7] **Alex Morales**, Nupoor Gandhi, Man-pui Sally Chan, Sophie Lohmann, Travis Sanchez, Kathleen A Brady, Lyle Ungar, Dolores Albarracín, and ChengXiang Zhai. Multi-attribute topic feature construction for social media-based prediction. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 1073–1078. IEEE, 2018.

[8] Sophie Lohmann, Benjamin X White, Zhen Zuo, Man-Pui Sally Chan, **Alex Morales**, Bo Li, Chengxiang Zhai, and Dolores Albarracín. Hiv messaging on twitter: an analysis of current practice and data-driven recommendations. *Aids*, 32(18):2799–2805, 2018.

[9] Man-pui Sally Chan, Sophie Lohmann, **Alex Morales**, Chengxiang Zhai, Lyle Ungar, David R Holtgrave, and Dolores Albarracín. An online risk index for the cross-sectional prediction of new hiv chlamydia, and gonorrhea diagnoses across us counties and across years. *AIDS and behavior*, 22(7):2322–2333, 2018.

[10] Ismini Lourentzou, **Alex Morales**, and ChengXiang Zhai. Text-based geolocation prediction of social media users with neural networks. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 696–705. IEEE, 2017.

[11] **Alex Morales** and ChengXiang Zhai. Identifying humor in reviews using background text sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 492–501, 2017.

[12] **Alex Morales** and ChengXiang Zhai. Forecasting research trends by mining literature findings. In *Proceedings of the 11th Annual ACM Richard Tapia Celebration of Diversity of Computing*. ACM, 2017.

[13] Huan Sun, **Alex Morales**, and Xifeng Yan. Synthetic review spamming and defense. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1088–1096. ACM, 2013.

[14] **Alex Morales**, Huan Sun, and Xifeng Yan. Synthetic review spamming and defense. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 155–156. ACM, 2013.