

Kate Silverstein

[linkedin: ksilverstein](#)
[k8si.github.io](#)
timezone: EST

Education

University of Massachusetts Amherst

M.S. in Computer Science.

Thesis: *Deep Emoticon Prediction for Weakly-Supervised Sentiment Analysis on Social Media Data*. Advised by Andrew McCallum.

May 2016

B.A. in Linguistics

May 2014

B.S. in Computer Science

May 2014

Work Experience

Senior ML Engineer. Established benchmark suite for evaluating language models on key downstream tasks, including toxicity detection, topic classification, and named entity recognition (NER). Pretrained multilingual language model on 1 billion tweets, improving performance across majority of benchmark tasks. Established evaluation protocol and baselines for video classification applied to ensuring brand safety for advertising.

Cortex NLP Signals Team
Twitter
June 2022 - January 2023

ML Engineer III. Led design, planning, & development of toolkit for scalable, multi-task training of Transformer-based text encoders with a team of 4 active contributors. Toolkit supported ML initiatives across 5 different projects. Led planning & development of multi-task model capable of generalizing well across a variety of downstream tasks for use in production chatbot platform. Contributed to research on self-supervised language model training with "anti-models" of negative data [Wick et al., 2020].

Machine Learning Research Group
Oracle Labs
August 2018 - May 2022

ML Engineer II. Developed and maintained production sentiment classification library. Implemented active learning loop for guiding crowdsourced data collection to reduce cost of continuous model re-training. Built "minimum viable" entity linking system for social media data.

Oracle Social Data Platform
June 2016 - August 2018

Research Intern. Developed novel learning and inference procedures for joint named entity recognition (NER), linking, and type resolution for social media data [Silverstein et al., 2015].

Machine Learning Research Group
Oracle Labs
Summer 2015
Advisor: Pallika Kanani

Research Assistant. Developed NER system for TAC KBP 2014 submission. Helped maintain NER components of FACTORIE library. Contributed to research on automatic extraction of academic paper headers using conditional random fields & character-level LSTMs. Contributed to research on dynamic feature selection for NER [Strubell et al., 2015].

Information Extraction & Synthesis
Lab (IESL)
School of Computer Science, UMass
May 2014 - June 2016
Advisor: Andrew McCallum

Other Proficiencies & Languages

English (native), Spanish (conversational), Latin (rusty)	Natural languages
Python, SQL, Java, bash, JavaScript, HTML/CSS, Scala, MATLAB/Octave	Programming languages
PyTorch, HuggingFace, TensorFlow, BigQuery, SciPy stack, NLTK, Praat, L ^A T _E X	Relevant libraries & software
git, Docker, Jenkins	DevOps

Patents

Pallika Haridas Kanani, Michael Louis Wick, Katherine Silverstein.
2019. *Named entity recognition and entity linking joint training*. U.S.
Patent 10410139.

Publications

Michael L Wick, Kate Silverstein, Adam Pocock, Jean-Baptiste Tristan,
and Mark Johnson. Detecting and Exorcising Statistical Demons from
Language Models with Anti-Models of Negative Data. *arxiv preprint*,
2020. URL <https://arxiv.org/pdf/2010.11855.pdf>

Kate Silverstein, Pallika Kanani, Michael Wick, Phillip Ogren, and
Stephen Green. Joint Training and Inference for Named Entity Ex-
traction and Linking on Twitter. 2015

Poster presented at the UMass Center
for Data Science Research Symposium.

Emma Strubell, Luke Vilnis, Kate Silverstein, and Andrew McCallum.
Learning dynamic feature selection for fast sequential prediction. In
Annual Meeting of the Association for Computational Linguistics, 2015

Outstanding paper award.

Benjamin Roth, Emma Strubell, Katherine Silverstein, and Andrew
Mccallum. Minimally Supervised Event Argument Extraction us-
ing Universal Schema. *Neural Information Processing Systems (NIPS)*
Workshop on Automatic Knowledge Base Construction (AKBC), pages 1–5,
2014a

Benjamin Roth, Emma Strubell, John Sullivan, Lakshmi Vikraman,
Katherine Silverstein, and Andrew McCallum. Universal Schema
for Slot-Filling, Cold-Start KBP and Event Argument Extraction:
UMassIESL at TAC KBP 2014. In *Text Analysis Conference (Knowledge
Base Population Track) '14 Workshop (TAC KBP)*, Gaithersburg, Mary-
land, USA, nov 2014b

Kate Silverstein and Kristine M Yu. Investigating tonal spaces using
an extension of VoiceSauce voice analysis software. *The Journal of the
Acoustical Society of America*, 134(5):4095, 2013. DOI: 10.1121/1.4830960

Poster presentation.