

# Michael W. Kearney, Ph.D.

Data Scientist [NLP](#) | [AI](#) | [ML](#)

mike.wayne.kearney@gmail.com

(816) 447-8650

[mikewk.com](#) [in/mikewk](#)  
[mkearney](#) [kearneymw](#)

## Professional Summary

Internationally recognized expert in Data Science, Natural Language, and Artificial Intelligence with over ten years of professional experience, strong theoretical and quantitative backgrounds, and a knack for finding and implementing innovative solutions to complex problems.

## Professional Experience

<b>AI Consultant</b>	Aware (client)	Michael Kearney AI Consulting LLC	<b>2023</b>
<ul style="list-style-type: none"><li>Delivered recommendations on use and development of generative AI and large language models</li><li>Described strengths and limitations in data sources and models and estimated project timelines</li><li>Drafted plans to implement/monitor human-in-the-loop AI labeling and offer multi-tier API access</li></ul>			
<b>Senior Data Scientist</b>	Language/Labs	<a href="#">Aware</a>	<b>2021-2023</b>
<ul style="list-style-type: none"><li>Refactored, optimized, and implemented best practices for large language models</li><li>Enabled multilingual training and converted pipelines from Airflow to Sagemaker Pipelines</li><li>Proposed and built information quality model to supercharge ML and generative AI</li><li>Curated external dataset containing 500 million heuristically labeled social messages</li><li>Created internal packaging for external data collection, data visualization, and sampling</li><li>Owned, improved, and expanded (to non-English) featured NLP models-sentiment &amp; toxic speech</li><li>Automated model reports on inputs, parameters, and metrics (precision, recall, f1, accuracy)</li><li>Used SQL and Athena to efficiently query customer data in DynamoDB, Athena, and Redshift</li><li>Frequently worked with datalakes, Apache Arrow columnar format, and unstructured JSON data</li><li>Mentored junior team members and often contributed tutorials and other documentation</li><li>Delivered public- and client-facing thematic analysis of workplace collaboration messages</li><li>Created labeling rule books, documented edge cases, and trained third party annotators</li></ul>			
<b>Assistant Professor</b>	<a href="#">Data Science &amp; Journalism</a>	University of Missouri	<b>2017-2021</b>
<ul style="list-style-type: none"><li>Developed and taught courses on social media data, networks, machine learning, and statistics</li><li>Mentored dozens of data science and journalism capstones and chaired two doctoral dissertations</li><li>Developed and reviewed open-source data science tools, including popular bot detection software</li><li>Served on various school and university committees related to data, education, and diversity</li></ul>			
<b>Data Consultant</b>	<a href="#">Research Methods Center</a>	University of Kansas	<b>2016-2017</b>
<ul style="list-style-type: none"><li>Delivered feedback on interdisciplinary research projects across university system</li><li>Led data science working groups on python, machine learning, and natural language processing</li><li>Contributed to project automating personalized analytics for students in a large school district</li></ul>			
<b>Instructor</b>	<a href="#">Communication</a>	University of Kansas	<b>2012-2016</b>
<ul style="list-style-type: none"><li>Created rtweet, a massively popular R package for collecting and analyzing Twitter data</li><li>Instructor of record for quantitative methods, public speaking, leadership, and persuasion</li><li>Updated and refined course materials for core quantitative research methods course</li></ul>			

## Tools

---

Languages	Python   R   SQL   Git   Docker   Bash   JavaScript
AWS Resources	S3   Sagemaker   ECR   Athena   Redshift   DynamoDB
Web Data Collection	REST/Stream APIs   requests   BS4   HTTP   Rvest
Dev-Ops	Azure   Jira   GitHub   Gitlab   Confluence   Trello
Deep Learning	PyTorch   TensorFlow   HuggingFace   Keras
Machine Learning	Scikit-learn   XGBoost   FastText   Tidymodels
Natural Language	Spacy   HuggingFace   NLTK   RegEx   Scikit-learn
Data Science	Polars   Pandas   Numpy   Tidymodels   Data.table
Data Visualization	Ggplot2   Ggvis   Plotly   Matplotlib   D3
Reporting	JupyterNB   Markdown   Knitr   Quarto   LaTeX

## Skills

---

Theory	Communication   Media   Interpersonal   Polling   Topic Mining
Research Methods	Experiments   Survey   Annotation   Algorithmic Labeling   Sampling
Artificial Intelligence	LLMs   Transformers   Human in the Loop   Reinforcement
Supervised Learning	Linear Models   Multilevel   Time Series   ANOVA   Structural Equation
Unsupervised Learning	Topic Modeling   Network Analysis   PCA   Factor Analysis   Similarity
Leadership	Project Management   Team Supervisor   Committee Member & Chair

## Education

---

<b>Ph.D.</b> (with Honors) <a href="#">Communication &amp; Behavioral Sciences Methodology</a>	University of Kansas	Lawrence, KS	<b>2017</b>
<b>M.A.</b> (with Honors) <a href="#">Communication Studies</a>	University of Kansas	Lawrence, KS	<b>2014</b>
<b>B.S.</b> (comprehensive) Socio-Political Communication	Missouri State University	Springfield, MO	<b>2012</b>

## Publications (Selected)

---

### Peer-Reviewed Articles

- Comparative approaches to misinformation, International Journal of Communication, 2021
- Exploring gender differences on Twitter with NLP, Journal of Language and Social Psychology, 2021
- Collecting and analyzing Twitter data, Journal of Open-Source Software, 2019
- Analyzing change in network polarization, New Media Society, 2019
- Two tests of social displacement through social media use, Information, Communication, & Society, 2018
- Interpersonal goals and political uses of Facebook, Communication Research Reports, 2017

### Popular Press Articles

- March Madness office talk on par with Vegas odds, PR Newswire, 2023
- Doubts cast over Elon Musk's Twitter bot claims, BBC News, 2022
- Using data science to uncover mystery author of Times' Op-ed, Reynolds Journalism Institute, 2018
- Trusting News project report, Trusting News, 2017

### Open-Source Software

- tweetbotornot2: Detecting Twitter bots, 100,000 downloads / 27 citations, 2019
- textfeatures: Extracts features from text, 43,000 downloads / 4 citations, 2017
- rtweet: Collecting Twitter data, 1,100,000 downloads / 491 citations, 2016