Dr. Michelle Morales

Data Science Professor

Franklin Square, New York
(516) 375-7775
Ms.morales3@gmail.com
https://michellemorales.github.io/

EDUCATION

2012 - 2017 PhD in Linguistics
CUNY Graduate Center
- Computational Linguistics certificate
- Dissertation: Multimodal Depression Detection: An Investigation of Features and Fusion Techniques for Automated Systems. Supervisors: Rivka Levitan, Andrew Rosenberg, Stefan Scherer, and Martin Chodorow

2010 - 2012 MS in TESOL
Queens College
- New York State of Education Teaching certificate

2007 - 2010 BA in Spanish Studies
Marist University

WORK & RESEARCH EXPERIENCE

Senior Machine Learning Engineer, Woebot Health
October 2022 - January 2025

- Led end-to-end ML lifecycle, including data manipulation and processing, model training, production deployments and post-deployment performance monitoring
- Built ML classifier internal-testing website using HTML, CSS & Javascript for pre-deployment readiness testing
- Built new and managed existing AI backend services using AWS, Python, and Javascript with full CI/CD pipelines in CircleCI, enabling automated unit tests and seamless deployment to AWS upon merges to GitHub. This setup ensured high code quality, reliable releases, and enhanced scalability for application performance
- Prompt engineered 20+ Generative AI prompts for a clinical trial app, used to study the effectiveness and safety of a fully generative relational agent for treating depression and anxiety

Principal Data Scientist, IBM (Research)
January 2021 - September 2022

- Served as liaison between IBM Consulting and IBM research, focused on identifying the most promising research assets which could be commercialized and leveraged by clients to solve business problems
- Served as research expert in technical sales conversations with IBM consulting clients, architecting possible AI based solutions. Team's highlighted client wins: Veteran Benefit Administration's mail automation ($40M), EY's document processing due diligence engine, and Standard Chartered's trade document processing engine
- Led effort to commercialize Computer Vision research asset; won first deal ($.5M) during first month of launch

Lead Data Scientist, IBM (Consulting)
July 2019 - January 2021

- Led team of data scientists in developing data pipeline and analytics using AWS cloud infrastructure, providing insights on COVID-19 case surge for RIDOH policy team
- Built LSTM based times series forecasting models using Python for a US pharmaceutical company, demonstrating a 25% increase in accuracy
- Built ETL pipeline, using Python and SQL, for combining internal inventory data with weather data to create time series dataset for forecasting product demand for large US retailer
- Presented in sales conversations explanations and value propositions of ML concepts to non-tech stakeholders / clients

Senior Data Scientist, IBM (Chief Analytics Office)
June 2017 - June 2019

- Built data pipeline for scraping publicly available IBM product reviews from online sites
- Leveraged an unsupervised algorithm for clustering product reviews to understand key themes
- Developed a novel patented competitive metric for understanding how IBM's products rank against competition, used internally by 10,000 IBMers

Research Intern, USC Institute for Creative Technologies
May 2016 - August 2016

- Built feature engineering pipeline in order to extract linguistically motivated features from text data to be used in multiple downstream ML systems, including depression detection systems as well as automated psychometric instruments for adult attachment

Web Developer, CUNY Graduate Center
August 2016 - December 2017

- Developed and maintained responsive website for the Futures Initiative using HTML, CSS, and Wordpress technologies
- Troubleshooted and resolved technical issues

Undergraduate Student Advisor, CUNY Graduate Center
August 2012 - August 2016

- Advised undergraduate CUNY students during their graduate school application process

TEACHING EXPERIENCE

Adjunct Assistant Professor, Farmingdale State College
January 2025 - Present

- Taught undergraduate courses for the Science, Technology and Society program. Courses include Scientific Thinking (Spring and Fall 2025), Data & Society (Fall 2025), and AI & Society (new course in development)

Associate Adjunct, Columbia University
June 2020 - Present

- Served as associate in the Data Science masters program in the School of Professional studies. Courses include: (1) Applied Text & Natural Analytics course and AI & User interfaces course

Teaching Assistant, CUNY Graduate Center
September 2014 - December 2014

- Served as teaching assistant for graduate level linguistics courses

Adult ENL Teacher, Great Neck Public School
June 2012 - August 2012

- Taught English as a new language classes to adult students

K-12 ENL Teacher, New York City Department of Education
September 2011 - June 2012

- Co-taught English as a new language classes in K-12 settings

SKILLS

- **Spoken Languages**: English and Spanish
- **Programming Languages:** Python, SQL, HTML, CSS and Javascript
- **Data / ML:** Anthropic API, BigQuery, DynamoDB, HuggingFace, Jupyter Notebooks, LangGraph/LangChain, MongoDB, Numpy, OpenAI API, Pandas, PyTorch, Sagemaker, Scikit-learn, and Spark
- **Cloud:** AWS and GCP
- **Monitoring / Automation:** Cloudwatch, CircleCI, Serverless, and Terraform

PATENTS

2022, Schoeffer, Jakob; **Morales, Michelle**; Sarvestani, Amir Sabet; Strategic planning using deep learning, US Patent App. 17/091,132

2021, **Morales, Michelle**; Usmani, Sheema; Srivastava, Biplav; Bernagozzi, Mariana Cecilia; Sarvestani, Amir Sabet; Klippel, Lee Case; Computerized competitiveness analysis, US Patent App. 16/564,252

2020, Weldemariam, Komminist; Speakman, Skyler; Goldberg, Itzhack; Kwatra, Shikhar; **Morales, Michelle**; Context enabled sender communication awareness alert, US Patent App. 16/257,583

PUBLICATIONS

Journal Articles

2017, Parra, F; Miljkovitch, Raphaële; Persiaux, Gwenaelle; **Morales, M**; Scherer, S; The multimodal study of adult attachment. Developing the Biometric Attachment Test, Journal of Medical Internet Research

Magazine Articles

2021, Huang, Yufeng; Bernagozzi, Mariana; **Morales, Michelle**; Usmani, Sheema; Srivastava, Biplav; Mullins, Michelle; Clarity 2.0: Improved assessment of product competitiveness from online content, AI Magazine

Conference Proceedings

2021, **Morales, Michelle**; Dey, Prajjalita; Kohli, Kriti; A comparison of simple vs. complex models for suicide risk assessment, Proceedings of the seventh workshop on computational linguistics and clinical psychology: Improving access

2020, Usmani, Sheema; Bernagozzi, Mariana; Huang, Yufeng; **Morales, Michelle**; Sarvestani, Amir Sabet; Srivastava, Biplav; Data-Driven Ranking and Visualization of Products by Competitiveness, Proceedings of the AAAI Conference on Artificial Intelligence

2019, **Morales, Michelle**; Dey, Prajjalita; Theisen, Thomas; Belitz, Danny; Chernova, Natalia; An investigation of deep learning systems for suicide risk assessment, Proceedings of the sixth workshop on computational linguistics and clinical psychology

2019, Loukina, Anastassia; **Morales, Michelle**; Kumar, Rohit; Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Industry Papers)

2018, **Morales, Michelle**; Scherer, Stefan; Levitan, Rivka; A linguistically-informed fusion approach for multimodal depression detection, Proceedings of the fifth workshop on computational linguistics and clinical psychology: from keyboard to clinic

2017, **Morales, Michelle**; Scherer, Stefan; Levitan, Rivka; A cross-modal review of indicators for depression detection systems, Proceedings of the fourth workshop on computational linguistics and clinical psychology—From linguistic signal to clinical reality

2017, **Morales, Michelle Renee**; Scherer, Stefan; Levitan, Rivka; OpenMM: An Open-Source Multimodal Feature Extraction Tool., Proceedings of Interspeech Conference

2016, **Morales, Michelle Renee**; Levitan, Rivka; Mitigating confounding factors in depression detection using an unsupervised clustering approach, Computing and Mental Health Workshop (CHI), San Jose, CA, USA

2016, **Morales, Michelle Renee**; Levitan, Rivka; Speech vs. text: A comparative analysis of features for depression detection systems, 2016 IEEE spoken language technology workshop (SLT)

2016, **Morales, Michelle**; Using Linguistic Knowledge to Automatically Learn Monotonicity Properties, Studies in the Linguistic Sciences, University of Illinois at Urbana-Champaign.

2015, An, Guozhen; Brizan, David Guy; Ma, Min; **Morales, Michelle;** Syed, Ali Raza; Rosenberg, Andrew; Automatic recognition of unified parkinson's disease rating from speech with acoustic, i-vector and phonotactic features., Proceedings of Interspeech Conference

Dissertation

2018, **Morales, Michelle**; Multimodal depression detection: An investigation of features and fusion techniques for automated systems, City University of New York

INVITED TALKS

- **2023 -** Black Girls Code presentation on becoming an ML Engineer
- **2017 -** Codeland presentation (video) on open source tools for Mental Health

GRANTS & FELLOWSHIPS

- **2016 - 2017**, Futures Initiative Fellowship, CUNY Graduate Center
- **2016 -** Provost's Digital Innovation Training Grant, CUNY Graduate Center
- **2016 -** Provost's Pre-dissertation Summer Research Grant, CUNY Graduate Center
- **2016 -** Doctoral Student Research Grant, CUNY Graduate Center
- **2012 - 2016 -** Presidential MAGNET Fellowship, CUNY Graduate Center