JOYSTON MENEZES

Los Angeles, CA | +1 (213) 519-8504 | menezesjoyston@gmail.com | joystonmenez.info
LinkedIn | GitHub | Google Scholar

RESEARCH INTERESTS

Health communication and health misinformation; exposure to harmful content on algorithmically curated platforms; digital mental health; measurement validity in computational social science.

EDUCATION

University of Southern California Jan 2023 – Dec 2024
Master of Science in Computer Science, Viterbi School of Engineering Los Angeles, CA

- GPA: 3.71/4.00. Degree conferred December 18, 2024.
- Coursework: Applied Natural Language Processing; Information Retrieval and Web Search Engines; Database Systems; Analysis of Algorithms; Web Technologies; Multimedia Systems Design.
- Honor: Order of the Laurel and the Palm (2025), USC's highest honor for graduating students, recognizing leadership, scholarship, and service.

NMAM Institute of Technology Aug 2017 – Aug 2021
Bachelor of Engineering in Computer Science Nitte, India

- Thesis: Detection and Visualization of Bone Abnormalities on X-ray Images Using Convolutional Neural Networks (Best Project of the University, 2021).
- Honors: Best Outgoing Student of the College and of the Computer Science Department; Outstanding Achievement Award (2021).
- Leadership: Student President; Branch Captain representing the Computer Science Department in technical and non-technical events.

PUBLICATIONS

Conference Papers and Proceedings

1. Pinto, G., Bickham, C., Salkar, T., **Menezes, J.**, Luceri, L., & Ferrara, E. (2025). Tracking the 2024 US Presidential Election Chatter on TikTok: A Public Multimodal Dataset. In *Companion Proceedings of the ACM Web Conference 2025 (WWW '25 Companion)*, 773–776. [DOI] [arXiv] [Dataset]
2. Shetty, S., Shetty, S., & **Menezes, J.** (2025). Augmenting Large Language Models with Dense Retrieval Techniques for Enhanced QA Performance. In *2025 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*.
3. Shetty, S., Hegde, S., Shetty, S., Shetty, D., Sowmya, M. R., Miranda, R., Sequeira, F., & **Menezes, J.** (2021). Deep Learning Photograph Caption Generator. In D. P. Shetty & S. Shetty (Eds.), *Recent Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2020* (Advances in Intelligent Systems and Computing, Vol. 1386). Springer. [Proceedings]

Book Chapter

1. Rodrigues, A. P., **Menezes, J.**, Fernandes, R., Aishwarya, Chiplunkar, N. N., & Padmanabha, V. (2021). Identification and Classification of Paddy Crop Diseases Using Big Data Machine Learning Techniques. In A. K. Tyagi (Ed.), *Data Science and Data Analytics: Opportunities and Challenges* (pp. 47–64). Chapman and Hall/CRC. [DOI]

Acknowledged Contributions

1. Pessianzadeh, A., Jamie, P., Sultana, N., Himmelstein, G., Zektser, Y., Ganz, P. A., Hosseinmardi, H., Ghasemian, A., & Rezapour, R. (2026). Beyond Binary Detection: A Multi-Dimensional Taxonomy of Cancer Misinformation on Reddit. In *Findings of the Association for Computational Linguistics: EMNLP 2026*, 40392–40419. Acknowledged for assistance and contributions to the work. [ACL Anthology] [arXiv]

RESEARCH EXPERIENCE

UC Davis Department of Communication and University of Warsaw Aug 2026 – Present
Research Collaborator Remote
PI: Prof. Magdalena Wojcieszak. Collaborators: Erin Wertz and Subhayan Mukerjee.

- Assessed the feasibility of cross-platform sockpuppet audits of recommendation, search, and AI search features across nine platforms (YouTube, X, TikTok, Instagram, Threads, Bluesky, Truth Social, Gab, GETTR), informing the team's audit planning.

- Built a celebrity-mention matching pipeline over 3.32 million political news posts and 29,999,224 Facebook posts, resolving canonical names from LLM-generated annotations and auditing ambiguous single-token names; matched 1,112,649 posts, reducing the team's political-classifier workload 27-fold.
- Built an outlet typology and per-celebrity concentration measures for a study of how celebrities' political engagement relates to news engagement, and identified 94,349 duplicate records in the source corpus.
- Fit a BERTopic model on a GPU cluster and validated it with a word-intrusion test deployed as a web-based coding task: three human coders identified the intruder word 93.5% of the time on average, with 86 to 92% inter-coder agreement.

OASIS Lab, Department of Communication, UCLA Feb 2026 – Present
Research Assistant Los Angeles, CA
Advisors: Prof. Homa Hosseinmardi and Prof. Amir Ghasemian

- Built an LLM classification pipeline (Python, OpenAI GPT-4o-mini) for a computational study of health misinformation on TikTok, using zero-shot structured outputs to enforce a six-dimension annotation taxonomy, and analyzed the distribution of high-risk medical claims across taxonomy dimensions.
- Contributed to a multi-dimensional taxonomy study of cancer misinformation on Reddit (Findings of EMNLP 2026), acknowledged by the authors for assistance and contributions.
- Implemented eight evaluation metrics comparing Gemini-generated replies with Reddit community replies to cancer misinformation (12,663 responses across 380 posts), including EPITOME empathy classifiers, readability, and an open-source LLM judge (Qwen 2.5) scoring harm and actionability.
- Audited metric validity: showed that a single empathy score hides a split (the model expresses warmth but rarely communicates understanding), rejected a toxicity classifier that flagged correct medical warnings as unsafe, and dropped judge scales that carried no information.
- Built a complete census of children's YouTube channels with more than one million subscribers: 103,066 videos, verified against each channel's reported upload count, with full metadata, a 17,517-video stratified sample, and transcript collection.

HUMANS Lab, USC Information Sciences Institute Aug 2024 – Jan 2026
Research Assistant Marina del Rey, CA
Supervisors: Prof. Luca Luceri and Prof. Emilio Ferrara

- Built scraping and preprocessing pipelines (Python, Pandas, Apache Spark) covering more than one million TikTok and X posts, plus ten million comments, on the 2024 U.S. presidential election, contributing to a public dataset of 3.14 million TikTok videos (WWW '25 Companion).
- Applied BERTopic topic modeling and VADER sentiment analysis to consolidate more than 700 fragmented topics into 10 coherent themes, and analyzed engagement and sentiment patterns across them.

Lerman Lab, USC Information Sciences Institute Aug 2024 – May 2025
Research Assistant Marina del Rey, CA
Supervisor: Prof. Kristina Lerman

- Co-developed a hierarchical taxonomy distinguishing harmful from recovery-oriented content on muscle dysmorphia ("Bigorexia") and helped build LLaMA- and Gemini-based classifiers to apply it to TikTok video data at scale, reaching 92% precision.
- Led collection and preprocessing of millions of TikTok videos and cross-platform data, and applied sentiment-aware topic modeling and multimodal embeddings to characterize how harm and recovery narratives co-occur.

NMAM Institute of Technology Jun 2020 – Jul 2021
Research Assistant Nitte, India

- Developed machine learning models (Decision Tree, Random Forest, SVM, Naive Bayes) to classify six common paddy crop diseases from leaf images; the work was published as a CRC Press book chapter.
- Implemented image preprocessing, Fuzzy C-Means segmentation on Apache Spark, and feature extraction, and benchmarked the speed-up and scale-up of Spark-based against Hadoop-based clustering.

TEACHING

University of Southern California Aug 2023 – May 2024
Graduate Course Producer, CSCI 585 Database Systems Los Angeles, CA

- Supported more than 600 students over two consecutive semesters under Prof. Saty Raghavachary.
- Held office hours, prepared course materials, and helped students work through assignments on data modeling, SQL, NoSQL, and MapReduce.

PROFESSIONAL EXPERIENCE

Center for Homelessness, Housing, and Health Equity Research, USC Jan 2023 – Present
Software Developer Los Angeles, CA

- Developed five Android applications for field data collection, including the DemoGraphic Count app supporting the Greater Los Angeles Homeless Count.
- Built, hosted, and maintained websites for the Center's projects, including an Olympics homelessness site, Housing LA, and the Homeless Count site, averaging 12,000 to 15,000 visitors per week.
- Produced GIS and spatial visualizations of homelessness patterns (ArcGIS, GeoJSON) and automated data ingestion with Python scripts.

Alphonso Consultancy Aug 2021 – Nov 2022
Website Developer and Maintenance Engineer Udupi, India

- Built responsive interfaces (Angular, JavaScript) with SQL and PHP back ends, improving operational efficiency by 40%.

Quality Theorem Jan 2020 – Aug 2020
Project Lead and Intern Bangalore, India

- Led development of an online examination platform (Node.js, MongoDB, REST APIs, Docker).

PATENT APPLICATIONS

Indian Patent Office, published applications; co-inventor. [Nitte patent register] [Patent Office Journal]

1. Automated Brain Tumour Diagnosis System and Method Thereof. No. 202441089819 (2024).
2. Method for Analysing Public Sentiment on YouTube Comments. No. 202441065225 (2024).
3. An Automated Extraction and Summarization System. No. 202441065150 (2024).
4. An E-Commerce Product Comparison System. No. 202441065313 (2024).

HONORS AND AWARDS

- National Overseas Scholarship, Government of Karnataka, 2024
- Top Data Collector and Surveyor, USC Suzanne Dworak-Peck School of Social Work, 2023
- Student exchange program representative for the Computer Science Department, Ritsumeikan University, Japan, 2020
- Representative of India, Asian Conference of Young Catholic Students, Indonesia, 2019

SERVICE AND COMMUNITY ENGAGEMENT

Greater Los Angeles Homeless Count 2023 – Present
Team Lead Los Angeles, CA

- Led the DemoGraphic Count initiative, coordinating data collection across Los Angeles to support regional homelessness policy and planning.

Housing and Water Access Projects 2024
Volunteer Mexico and Colombia

- Tijuana, Mexico (Nov 2024): built homes with Esperanza International. Colombia (Mar 2024): helped build the first water tank for an Indigenous Wayuu community.

National Service Scheme, Better India Mission Jul 2017 – Dec 2021
Volunteer Leader Karnataka, India

- Implemented household waste management in villages; started the Piggy Bank project supporting homeless individuals and people living with HIV.

SKILLS

Programming: Python, JavaScript and Node.js, TypeScript, Java, C++, SQL, NoSQL

Methods: NLP and topic modeling (BERTopic), LLM-based classification (LLaMA, Gemini, GPT) with structured outputs, sentiment analysis, large-scale data collection and entity matching, social media data analysis

Tools: Pandas, PyTorch, TensorFlow, scikit-learn, DuckDB, Apache Spark, Git, Docker, ArcGIS

Languages: English, Hindi, Kannada, Konkani, Tulu