BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Yang Liu

eRA COMMONS USER NAME (credential, e.g., agency login):

POSITION TITLE: Graduate Teaching Assistant

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
North Carolina A&T State University, Greensboro, NC, USA	Ph.D.	05/2024	Computer Science
North Carolina A&T State University, Greensboro, NC, USA	M.S.	07/2020	Computer Science
North Carolina A&T State University, Greensboro, NC, USA	B.S.	05/2018	Electrical Engineering

A. Personal Statement

I am a Ph.D. candidate in the Computer Science Department at North Carolina Agricultural and Technical State University. I have been working as a graduate research assistant in Human-Centered AI (HCAI) Lab and Quantum Computing Research Center <QC|RC>. My research area covers machine learning, natural language processing, and Quantum Computing.

My master's research focused on understanding how language embedding plays an important role in sentiment analysis. Informed by my master's work, my Ph.D. work contributes to knowledge extraction from noisy social media data for public health surveillance. In general, my focus is to integrate social media sources with clinical surveillance systems effectively. In that way, we can provide early warnings for pandemics and make evidence-based decision making. From June 2021, I joined the Quantum Computing Research Center as a quantum information science and engineering (QISE) scholar. I have started studying state-of-the-art quantum computing technologies for machine learning and NLP.

I completed a journal publication titled "Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning" in June 2021. This study aims to explore "how useful is Reddit social media platform to surveil COVID-19 pandemic?" and "how do people's concerns/behaviors change over the course of COVID-19 pandemic in North Carolina?". We collected COVID-19 related data from 18 subreddits of North Carolina from March to August 2020. Next, we applied methods from natural language processing and machine learning to analyze collected Reddit posts using feature engineering, topic modeling, custom named-entity recognition (NER), and BERT-based (Bidirectional Encoder Representations from Transformers) sentence clustering. The study shows the utility of NLP methods in discovering the change of the public's concerns/behaviors over the course of COVID-19 pandemic in NC using Reddit data. Moreover, the results show that social media data can be utilized to surveil the epidemic situation in a specific community.

During the past summer of 2020, 2021, and 2022, I was selected as a summer camp mentor. The mentor group consisted of two undergraduate students from National Science Foundation (NSF) Research Experiences for Undergraduates (REU) Program and three graduate students from HCAI Lab. The primary objective of the program is to provide students an opportunity to conduct and present research with a diverse group of peers. It is targeted for undergraduate students who are interested in Machine Learning and NLP. Based on my research, I was involved with planning and executing the REU program, trained two students on variety topics within the field of NLP for ten weeks, and guided students to produce an oral presentation, a poster, and a manuscript after the summer program. I was able to showcase my leadership skills, but it also allowed me to be innovative in a diverse team environment. I researched, designed, lectured, organized, collaborated, and learned of many latest NLP topics. I was selected as a teaching assistant in QISE Undergraduate Scholar Program in Fall 2022.

Based on the contributions of my publications in the NLP area, I am very keen to contribute to the NC A&T Scientific Research Preparatory Program (SRPP) with other researchers in the field. Given my diverse research experience, I will be able to take full advantage of this program. It also will keep me updated on new findings that have taken place in my research area.

B. Positions, Scientific Appointments, and Honors

Positions:

North Carolina A&T State University
Ph.D. Candidate in Computer Science — Exp. May. 2024
M.S. in Computer Science — Jul. 2020
B.S. in Electrical Engineering — May. 2018

Scientific Appointments:

North Carolina A&T State University, Computer Science Department
Graduate Research and Trashing Assistant — Aug. 2020 - Present
IBM-HBCU Quantum Center
Quantum Information Science and Engineering (QISE) Scholar — Jun. 2021 - Present
National Science Foundation (NSF) Research Experiences for Undergraduates (REU) Program in Data Science
Research Mentor — Summer 2020 & 21 & 22

Honors:

- 2022 NC A&T Rising graduate Scholar — Apr. 2022
- Quantum Information Science and Engineering (QISE) Scholar — Jun. 2021 - Present
- Graduate Teaching Assistantship Winner — Aug. 2020 - May 2021
- Good Academic Standing for the Spring 2021 and Fall 2020 semesters — Aug. 2020 - May 2021
- Graduate Research Assistantship Winner — Aug. 2018 - May 2020
- Undergraduate Research Graduation Cords Award — May 2018
- Recognition of Outstanding Academic Achievement — May 2018
- Chancellor's List for the Spring 2018 — May 2018
- Dean's List for the Fall 2017, Spring 2017, Fall 2016 — Dec. 2016 - May 2017
- School-level 3rd Class Scholarship — May 2016

C. Contributions to Science

Served as a Reviewer for

- Conference Articles
  - The 13th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2022)
  - 2022 International Conference on Big Data Mining and Information Processing (BDMIP 2022)
- Journal Articles
  - The Journal of Medical Internet Research (JMIR) Publications
- Contest Activity
  - Host, 2022 Triad Programming Contest Feb. - Apr. 2022

Publications:

[1] **Liu, Y.**, Yue, Z., & Anwar, M. (2022, November). Monkeypox At-a-glance from Google Trends and Reddit. In 2022 IEEE/ACM international conference on Connected Health: Applications, Systems and Engineering Technologies (*CHASE*). (Accepted)

[2] **Liu, Y.**, & Anwar, M. (2022, September). Learning Programming in Social Media: An NLP-powered Reddit Study. In 2022 *International Conference on Transdisciplinary AI (TransAI*). IEEE. (Accepted).

[3] **Liu, Y.**, Whitfield, C., Anwar, M, et al. (2021). Monitoring COVID-19 pandemic through the lens of social media using natural language processing and machine learning. *Health Information Science and Systems*, 9(1), 1-16.

[4] Whitfield, C., **Liu, Y.**, & Anwar, M. (2021, August). Surveillance of COVID-19 pandemic using social media: a reddit study in North Carolina. In *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 1-8).

[5] Almousa, M., **Liu, Y.**, Zhang, T., & Anwar, M. (2021, July). A Study on Online Businesses' Commitment to Consumer Privacy. In *International Conference on Human-Computer Interaction* (pp. 391-402). Springer, Cham.

[6] **Liu, Y**. (2020). *A Comparative Study of Vector Space Language Models for Sentiment Analysis Using Reddit Data* (Master Thesis, North Carolina Agricultural and Technical State University).

[7] Vaughn, A., Biocco, P., **Liu, Y**., & Anwar, M. (2018, July). Activity detection and analysis using smartphone sensors. In *2018 IEEE International Conference on Information Reuse and Integration (IRI)* (pp. 102-107). IEEE.