Carnegie Mellon University

School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

cs.cmu.edu

May 30, 2023

Dear Committee Members,

I am pleased to recommend Prof. Mona Diab for the King Salman Prize for Arabic Language Technologies dedicated to acknowledging personal achievement in the field of Arabic Language technologies. Prof. Diab is joining CMU this summer as Director of CMU's Language Technology Institute, a world-renowned organization in research and education in all aspects of language technologies.

Prof. Diab is recognized as a top researcher in the field of Natural Language Processing (NLP), Applied Machine Learning, and Artificial Intelligence. Her research currently revolves around Responsible AI, Low Resource Language Processing, Multilinguality and crosslinguality. She uses Arabic NLP as a microcosm of the field at the intersection of all these thrusts. Over her 20 year career, Dr. Diab made significant contributions to Arabic NLP and Technologies. She pioneered research in Arabic variants processing such as Sentiment analysis, social media analysis, syntactic and semantic processing, linguistic code switching, etc. Her work, together with her collaborators, was not only about techniques used for English then transferring to Arabic, but rather devising new techniques to handle the complexity of Arabic both as a language family with myriad variants from classical to dialectal vernaculars and its morpho-syntactic structures relative to English. Her work pushed the field of Arabic processing forward but also pushed the boundaries of NLP as a science. Fall 2023 will mark the first year the community is having a dedicated Arabic NLP conference under the auspices of the foremost NLP community, ACL.

She is one of the leading figures who established modern Arabic NLP as a field within NLP/AI. She is considered one of the references in the field with over 150 research publications, theses, and technical reports covering Arabic NLP, establishing standards and research directions. She also built practical systems spanning both academia, government, and industry from prototypes to systems deployed for the US government (Madamira, CodACT, Colaba), industrial level systems (Google using Madamira, powering Meta Language Identification, Meta Machine Translation, Meta Arabic Hate Speech Labeling and Modeling). One of the highlights of Mona's research has been about enhancing readability focusing on simplification and partial diacritization. Mona pioneered work on linguistic code switching processing. Despite developing generalized models, her prism has always been through the Arabic language and its variants, again challenging established NLP paradigms. Mona's research on Arabic spans creating resources and models for classical Arabic (Semantic structures of the holy Quran), Modern Standard Arabic, and Social media informal genres, aka Arabic dialects.

Carnegie Mellon University

School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

cs.cmu.edu

Research is only one of the dimensions of Dr. Diab's contributions to Arabic NLP. Dr. Diab has spent the better part of her career mentoring and growing Arabic NLP talent both in the USA and locally in the Arab world. She (co-)supervised and graduated several Arabic NLP researchers many of whom are currently in their home countries serving as academic faculty and industry researchers. She also made a concerted effort to publish in Arab world conferences so that she would guarantee maximal exposure of local talent to the latest NLP/AI science, taking it to their doorsteps. She has been evangelizing the democratization of science through diversity and inclusion with the mission of propelling Arabic scientific contribution harnessing Arabic NLP spurring a new digital information and knowledge based economy. Dr. Diab was also privileged to bring Arabic NLP to the forefront of international top scientific venues in her conference keynote addresses and invited talks at world renown institutions globally, thereby educating general NLP practitioners on the nuances of the Arabic language and culture while highlighting the challenges posed by the complexity of the language from a technological perspective.

Sincerely,

Martial Hebert
Dean and University Professor
School of Computer Science
Carnegie Mellon University