Responsibility in AI Systems & Experiences (RAISE) at the University of Washington presents:

Olfa Nasraoui

Explainability by Design

Friday, May 6, 2022, 9-10am PT

Join: https://washington.zoom.us/j/94636255672

Abstract: At its core, AI is enabled by advanced Machine Learning (ML) models that are now being used increasingly to enable decision making in many sectors, ranging from e-commerce to health, education, justice, and criminal investigation. Hence, these algorithmic models are starting to directly interact with and affect the daily decisions of more and more human beings. In particular many models are black box models that make predictions without any justification to the user. Without any mechanism to allow humans to understand and question the reasons behind them, Black Box predictions lack justifiability and transparency. In addition, they cannot be scrutinized for possible mistakes and biases. Therefore, designing explainable machine learning models that facilitate conveying the reasoning behind their predictions, is of great importance. Yet, one main challenge in designing ML models is mitigating the trade-off between an explainable technique with moderate prediction accuracy and a more accurate technique with no explainable predictions. This talk will focus on recommender systems, a special family of Machine Learning models that is remarkably interactive with humans, and will present recent research, at the Knowledge Discovery & Web Mining Lab, in building explainability into a selection of state of the art Black Box recommender systems.

In particular, we present a framework for building machine learning models that are taught to make explainable predictions for a special case of explanations criteria that are designed according to the application domain. First we make the case for explainability by design as an attempt to fulfill the user's need and definition of an explanation for a particular domain and context.
While designing explanations and explainability, we also raise the need for fair explanations since an explanation can itself be prone to bias and unfairness.
Some motivations for explainability by design include (1) convenience of human design of the explainability and fairness criteria to satisfy one's conceived notion of what is explainable, even those that may be subjective and complex to formulate, (2) ability to formulate the machine learning problem in such a way to satisfy diverse desiderata that include the traditional learning task and the newly designed explainability and fairness criteria, (3) ability to verify that a machine learned model satisfies criteria that have been designed, and (4) ability to embed Explainability by design into various machine learning mechanisms due to the flexibility of the approach, for instance via modification of the loss function.

Bio: **Olfa Nasraoui** is Professor of Computer Science & Engineering, endowed Chair of e-commerce, and the founding director of the Knowledge Discovery & Web Mining Lab, in the Speed School of Engineering at University of Louisville. She conducts research in machine learning, AI, and data science, in particular web mining, information retrieval and recommender systems; and fairness and explainability in AI. She is a National Science Foundation CAREER award winner and twice winner of Best Paper Awards in the research area of machine learning and Fair AI. The first Best Paper Award is in Theoretical Developments in Computational Intelligence at Artificial Neural Network In Engineering (ANNIE 2001), on robust clustering algorithms for web usage mining. The second Best Paper Award is at KDIR 2018, for research on modeling and studying the impact of algorithmic bias in machine learning and recommender systems. She has served as Primary Investigator or Co-Investigator for over 20 research grant projects, funded from diverse sources, including the National Science Foundation and NASA. Olfa leads as PI the NSF funded ATHENA ADVANCE initiative for faculty equity at University of Louisville. She serves as Associate Editor for the Recommender Systems section of Frontiers in Big Data, and on the Editorial board of the International Journal of Machine Learning and Computing. She has also served as Associate Editor for IEEE Access and guest editor for Data Mining and Knowledge Discovery. She has served on the organizing and program committees of several conferences and workshops, including co-organizing the premier series of workshops on Web Mining, WebKDD 2004-2008, as part of ACM-KDD. She has also served as Program Committee Vice-Chair, Track Chair, or Senior Program Committee member for several data mining conferences including ACM RecSys, KDD, AAAI, IJCAI, ICDM, SDM, and CIKM. She is a member of ACM, ACM SIG-KDD, AAAS, and a senior member of IEEE.

RAISE is a UW-wide group of students and faculty interested in the broad space of responsibility in AI, trustworthy machine learning, human-centered computing and data science. As part of this group, our mission is to engage in scholarly, educational, and outreach activities that lead to foundational research in these areas.
https://www.raise.uw.edu

UNIVERSITY of WASHINGTON

RAISE
RESPONSIBILITY IN AI SYSTEMS & EXPERIENCES