Understanding Human-Machine Decisions About Race and Gender When Developing Computer Vision Models

I research marginalized identities through the "eyes" of computer vision (CV) models to make computer vision more human-centered, just, and equitable. At a granular level, I will contribute to growing interdisciplinary needs in algorithmic fairness between both human-computer interaction and machine learning. At a larger societal level, my research will provide key stakeholders—whether engineer or policymaker—opportunities for interdisciplinary scholars and practitioners to interrupt unjust identity represented in key areas of machine learning pipelines. My research will produce a holistic understanding of how human characteristics are operationalized and propagated in CV algorithms, towards developing actionable ways to improve the constructed nature of algorithmic identity—from the conception of an ML product to its inevitable impact on "users."

Aspects of human identity have long been embedded into our computerized systems. As we enter the age of big data and machine learning (ML)-based artificial intelligence (AI), computer systems are increasingly trained to recognize and respond to human characteristics. We already know there are often problems with how human identity is represented in data. These problems further cascade into ML-driven technologies, as human traits are embedded into algorithmic computer systems in new and unique ways, without "user" input. Contestability and control over how one's "data double"—the algorithmic construction of a human being—is created and used is increasingly difficult.

To investigate representations of human identity in ML, I focus on two types of identity that have experienced a history of marginalization and thus necessitate intervention: gender and race. I will examine gender and race through the lens of a particularly evocative ML domain in regards to identity representations in ML: *computer vision (CV)*, in particular *facial analysis (FA)* technologies. Facial analysis describes a whole host of face-based vision tasks, like *facial detection* and *facial recognition*.

Already, concern about FA stems from the use cases of CV-driven technologies, such as potentials for policing, social control, and surveillance by both state and corporate interests. People with minoritized racial and gender identities express unease about the notion of normative constructions of human identity encultured in CV. FA is particularly pertinent because it makes simple decisions about human identities based on visual data alone. Specifically, that human identities, like race and gender, are objective and simple to classify. Researchers are increasingly questioning *what* the results of racialized and gendered CV technologies are. However, few are asking *how* these systems are racializing and gendering people in the first place. The few researchers who have explored the "how" have examined components of CV technologies in isolation; for example, the outputs of classification, the academic literature, or the data publicly available for training the ML algorithms.

Yet CV technologies, whether used for FA or object detection, are not constructed by one individual in isolation, but by a series of culturally-situated individuals constructing a pipeline of systems. Such systems are constrained by economic, technical, and social limitations; for example, the economic limitations to constructing robust datasets, the technical limitations of ML being able to understand complex visual context, and the human limitations of developing diverse and inclusive labels.

Instead, I will explicitly examine FA as a set of multi-layered infrastructures. Building on the foundation of work I completed on FA and gender—where I examined gender classification in commercially available facial analysis technologies—I will contribute a holistic understanding of how human characteristics are operationalized and propagated by FA across multiple *layers* of human and computer actors: (1) at the model level, by analyzing the process of human annotation ; (2) at the annotation level, by deeply analyzing, both quantitatively and qualitatively, the work practices of human data annotators; and (3) at the development level, by studying the practices and constraints when developing computer vision systems, including how stakeholders make decisions about embedding identity characteristics into systems.

All these efforts will culminate in a deep understanding of where in the pipeline identity is embedded, how identity is understood by the model, annotators, and developers, and how the end users who interact with computer vision can then interpret that implementation. These understanding will allow designers, researchers, and engineers to intervene at key points of the pipeline. Targeted intervention can improve bias mitigation and ethical representation of marginalized identities.