Torbet McNeil, Research Statement

The interdisciplinary field of data science, which applies techniques from computer science and statistics to address questions across domains, has enjoyed recent considerable growth and interest. This emergence also extends to undergraduate education, whereby an increasing number of institutions now offer degree programs in data science. However, there is considerable variation in what the field actually entails, and, by extension, variation in how higher education institutions prepare students for data-intensive careers. Data science offers great hope for bringing the best knowledge available to address wicked problems facing the United States and world; yet, potential deficiencies in data science curricula may present challenges.

My dissertation investigates the training students receive when they pursue undergraduate data science degrees and organizational growth of programs in a segmented U.S. higher education marketplace. Higher education theories, such as academic capitalism and isomorphism, help elucidate what is driving differences between seminal data science curricular frameworks and offered curricula and adoption of degree programs. The two frameworks used suggest that considerable expertise is required by data science practitioners, and perhaps a four-year undergraduate degree program cannot cover all the necessary skills. However, higher education institutions should strive to appropriately brand their programs to minimize the gap between expectations and the training students receive.

Many higher education institutions approach data science education as a largely "skills-intensive," technical endeavor. However, such reductionist thinking of data science in "conventional" ways may have enormous consequences. Namely, the infatuation with "data," "science," and "intelligence" may lead to neglectful thinking of societal values and ethics. Critical curricular omissions potentially create a *Promethean* workforce prepared to use a variety of computational and statistical tools in socially inappropriate ways. Despite some leading scholars' cautions (Irizarry, 2020), formal undergraduate data science degree programs continue to emerge in a volatile national market often aggressively competing for students.

My research is published in *PeerJ Computer Science* (2021) and highlighted in a brief in *The Conversation* (2021) that was picked up by a diversity of outlets, including GovTech, The Next Web, NextGov, *Business Insider*, *University World News*, and the *Houston Chronicle*. I am currently working on two additional articles grounded in my dissertation research: "Comparative Analysis of Data Science Undergraduate Degree Programs" for *Policy Sciences* and "Investigating Institutional Characteristics of Adopters of Data Science Undergraduate Degree Programs Through Machine Learning" for *Higher Education*.

In my next project, I would like to expand on my work on what "good" data science is, including a focus on ethical practices and approaches to machine learning algorithms. Many programs have embraced data science education with a kind of inertial sensibility, i.e., it is inevitable that data science needs to be offered to students. Information and data affect so many areas of knowledge that higher education institutions feel they must assure students are "ready" for the world they are inheriting. Drawing from diverse data science educators and higher education administrators, I will conduct repeated surveys involving individual panel data collection, both cross-sectional and longitudinal, and carry out structured interviews.

Getting the first principles and educational offerings established are critical, especially as data science now encompasses so many fields. One's value outcomes can be skewered, and a broader set of norms can be ignored with incomplete or "partial" coverage of data science. For example, one can fail to ask the "right" policy questions. Topics to which I am contributing include how data science education should work in practice and how disciplines can coexist with each other in this space.