**The Data Science Ethos Lifecycle Framework: Ethical Impact for Meaningful Change**

Cesar Hernandez

School of Computing and Information, University of Pittsburgh

CMPINF 2140: Responsible Data Science

Dr. Matt Burton and Dr. Nora Mattern

2/24/2025

**The Data Science Ethos Lifecycle Framework: Ethical Impact for Meaningful Change**

The Data Science Ethos framework (referred to as the "Ethos Lifecycle") was developed to integrate ethical considerations with the overall data science process (Boenig-Liptsin, Tanweer, & Edmundson, 2022; Academic Data Science Alliance, 2022). Its key principles emphasize transparency, responsibility, and inclusivity in data practices. The framework encourages data scientists to consider human context and ethical consequences at each stage of their workflow, helping them foster responsible practices that protect and support the communities they engage with (Boenig-Liptsin, Tanweer, & Edmundson, 2022).

The Ethos Lifecycle comprises lenses, stages, and case studies. The lenses (Positionality, Sociotechnical Systems, Power, and Narratives) guide ethical analysis at each stage of a data science project. Positionality explores human diversity, Sociotechnical Systems examine the tech-society relationship, Power highlights asymmetries in agency, and Narratives focus on how stories influence change (Academic Data Science Alliance, 2022). The Ethos Lifecycle stages (Question, Discovery, Analysis, Modeling, Interpreting, and Sharing) represent key elements of the research process, from generating the research question to disseminating results. At each stage, the lenses help consider social impact and ethical issues. These lenses help evaluate a project through diverse human perspectives, technology impacts, behavioral influences, and future implications.

The final component, case studies, demonstrates how the framework applies to real-world scenarios (Academic Data Science Alliance, 2022). For example, using cell phones for data collection may be more costly and time-consuming than traditional census methods. By applying the four lenses during the Data Discovery stage, ethical questions arise: Positionality considers privacy, Sociotechnical Systems debates if the data represents individuals or families, Power assesses who controls the data and user awareness, and Narratives questions if cellphone data can genuinely answer the same questions as census data or offers new insights.

**Personal Resonance**

The Ethos Lifecycle has solidified my understanding of responsible data science in three ways: (1) every data point represents a person, (2) my responsibility extends beyond my current project, and 3) experiences influence the whole data science project.

            Initially, I saw data as just numbers, disconnected from the people behind it. In my current role, I manage a database of participants visiting our clinic for dementia-related concerns. I work with this data daily, sometimes without thinking about how it could help others or how complex it is. After learning about different frameworks, I'm reminded that each data point is a snapshot of someone's life, and it's my responsibility to respect and protect both the information and the integrity of my work. The funny thing is, I already feel this way, but the Ethos Lifecycle framework helps me keep it present throughout the project. It's like "treat others how you want to be treated," but for data: "treat the data of others how you would want yours to be treated."

            Specifically, the lenses in this framework have allowed me to be respectful and understanding at each stage because they examine the relationships and impact of a data science project. I am starting to see the patterns between data and society and data with technology, how those relationships influence one another, and their impact on the future. I like that I am more reflective about current and future problems, as it helps to be transparent and set better goals. I am constantly thinking about how stakeholders and collaboration influence the direction of goals and the overall structure of a project. I love that I am thinking of my assumptions and biases and the influence of power in formulating my projects, collecting data, and analyzing data.

            Positionality is the most interesting concept in the Ethos Lifecycle. It highlights how our uniqueness shapes every part of a data science project and challenges the idea that data and analysis are purely objective (Academic Data Science Alliance, 2022). It's made me realize how much my intentions in structuring a project matter. Positionality also ties into power dynamics—who collects, analyzes, and benefits from the data. Awareness of this makes me more intentional and accountable in my work, ensuring I respect the people behind the data, not just the numbers.

**Professional Application**

The Ethos Lifecycle would be incredibly helpful in my career as a data scientist because it would keep me grounded in ethical practices while considering the broader societal implications of my work. I am working on a project about social cognition changes in neurodegenerative diseases. Using all four lenses, I could implement the Ethos Lifecycle through the six stages. The following is an example of how I would implement this framework at various project stages.

***Positionality***

Usually, I clean my data efficiently by removing potential outliers that may skew it. But now, I may be removing data that could be meaningful. If I reflect on my assumptions, I can better decide what is needed during data cleaning.

***Power***

The population of interest is individuals with neurodegenerative diseases, who may not have much say in how their data is used. Understanding that power dynamic has helped me better advocate for more transparency in how we disseminate the results to patients and caregivers. I will work with my P.I. to determine how to share the results with those involved.

***Sociotechnical Systems***

In this project, I realized that some fields were collapsing categories for a particular variable, and more than 50% of participants were labeled "Other," so I worked with the other researchers to expand the options and create a write-in variable. Understanding the complexities of experience reminds me that sometimes, variables don't fully capture the patient's experience.

***Narratives***

Changes in social cognition happen for almost everyone, but the way the results of this project are conveyed is significant. Without context, people may become scared instead of focusing on early intervention or support. I would work with my P.I. to ensure that we frame the narrative as concentrating on support and early intervention and that changes in social cognition don't necessarily mean something is wrong.

**Critical Analysis**

            The framework is primarily reflective and conceptual but could benefit from a more structured approach with concrete steps. The guidelines can be abstract and overwhelming, so having specific questions for each lens at every stage would provide a clear pathway for users and ensure consistent application of ethical considerations. Since open-source lesson materials are currently in development, I'm excited to see how they incorporate these elements and develop a standard operating procedure (SOP).

            Another improvement could be adding a fourth component, which is auditing the processes. This new component would maintain ethical standards throughout and even after a project is finished. It would also ensure long-term accountability and help build a culture of ongoing ethical responsibility. Plus, it would allow us to address any unexpected issues during or after the project, keeping the framework relevant and effective over time.

            Adding domain-specific case studies may also open the framework to a broader audience. Currently, three cases are listed on the Academic Data Science Alliance's (2022) webpage: one focusing on population estimates using cell phone data, one on unfair tax practices, and the last on investigating piping infrastructure for water pipes. The more case studies there are, the easier it could be to generate questions about navigating the stages with input from the lenses. More case studies may show how applicable the framework is as it can be very abstract and idealistic. Collaborating with different sectors or industries could help achieve this.

**Conclusion**

                  The Data Science Ethos Lifecycle Framework resonates with me because it emphasizes the importance of ethical considerations and social implications at every project stage. It focuses on how data, people, and technology are interconnected while considering the present and future. At its core, it promotes transparency, accountability, privacy, fairness, and non-maleficence—key principles of responsible data science. It also encourages ongoing reflection on how your work impacts society and how your personal experiences might shape your decisions, leading to a more responsible, people-focused approach.

**References**

**Academic Data Science Alliance. (2022).** The Data Science Ethos. Retrieved February 24, 2025, from <https://ethos.academicdatascience.org/>

**Boenig-Liptsin, M., Tanweer, A., & Edmundson, A. (2022).** Data Science Ethos Lifecycle: Interplay of ethical thinking and data science practice. Journal of Statistics and Data Science Education, 30(3), 228–240. <https://doi.org/10.1080/26939169.2022.2089411>