Ethical Challenges in the Capstone Project

Tene Ariyo

Ethical Challenges and How They Were Addressed

1. Data Privacy and Consent

A major ethical consideration in this project was ensuring that any data used, especially if it involved students, teachers, or minors, respected privacy and consent standards. Because my *R for Everyone* curriculum is designed for middle school classrooms, I needed to make sure no identifiable or sensitive information from students would ever be stored, shared, or analyzed.

How I addressed it: - Used only simulated or publicly available datasets rather than collecting student-level data.
- Designed the *learnr* lessons to process data locally in RStudio Cloud or on the user's device without exporting responses externally.
- Removed any automatic data-logging features from quizzes or free-response questions to avoid saving identifiable information.

2. Informed Participation

Since this project may later involve classroom pilots, I considered what it would mean for teachers and students to participate voluntarily and knowingly.

How I addressed it: - Created a teacher validation rubric for expert review rather than using live student data in this phase.
- Included a short consent statement in the teacher feedback form clarifying that participation is voluntary, anonymous, and for educational research purposes only.

3. Algorithmic Bias and Representation

Teaching data science to younger audiences raises the risk of unintentionally reinforcing biased or incomplete views of data. For instance, datasets that overrepresent certain regions or demographics could lead students to draw skewed conclusions.

How I addressed it: - Added discussion prompts encouraging students to question data sources, collection methods, and who might be excluded.

- Included a short "ethics micro-prompt" in each lesson, for example: *"Whose data is missing here, and how might that affect the story this graph tells?"*

4. Accessibility and Digital Equity

Because not all schools or students have the same access to technology, the project needed to be designed with varying levels of connectivity and device availability in mind.

How I addressed it: - Structured the *learnr* lessons so they can be run offline in RStudio or RStudio Cloud using minimal computing resources.
- Provided printable summary worksheets and screenshots for teachers with limited device access.
- Planned for open-source release on shinyapps so that all materials remain free to use and adapt.

5. Pedagogical Transparency

Finally, I wanted to ensure that the project models ethical data literacy itself, students should learn not only *how* to analyze data but *why ethical reflection matters*.

How I addressed it: - Embedded reflection checkpoints after each coding task, asking students to consider reliability, and implications of their analyses.
- Used plain-language explanations to make ethics approachable rather than abstract or punitive.
- Documented all datasets, their origins, and transformations in reproducible RMarkdown cells.

Summary:
Throughout this capstone, ethical considerations guided both the *design* (how data are collected and represented) and the *delivery* (how learners interact with R safely and thoughtfully). The overarching goal was not just to avoid harm, but to teach ethical awareness as a core part of data science practice.