I used to consider data science and machine learning with respect to the development of the most accurate and efficient models, optimization of data processes, and effective delivery of actionable insights. Looking back at past work today, I find it important to incorporate data ethics and inequity considerations into every aspect of a project. As the field of data science continues to evolve, our modes of thought around ethical consequences must likewise do so.

Data ethics encompasses a wide range of issues, from privacy and consent to bias and fairness. For previous projects where machine learning algorithms, for example, have been developed, it would be important to have the processes through which data are collected very transparent and that the participants are well informed about their data usage. It would not only involve taking prior explicit consent but also ensuring data anonymization in order to protect individual privacy. I would also introduce rigid bias checks into the datasets. It would mean actually looking out for and correcting imbalances that would cause unfair or skewed outcomes. Suppose there were cases where some dataset overrepresents one demographic; I would do re-sampling or introduce fairness constraints to ensure the model does not project predictions which would further extend inequities that already exist.

I would also bake ethics into the deployment phase for machine learning projects. This involves much-needed continuous monitoring to identify unwanted consequences in model predictions that could stereotype or disadvantage any group. In practice, this may be achieved by constructing feedback loops whereby users can report perceived biases or inaccuracies; after analysis, that feedback could be used to tune the model.

I would also advocate for the development process being represented by a multitude of voices. This would involve various stakeholders in project planning and cultivating an enabling environment whereby concerns about equity and fairness are raised freely. That way, I would develop models that not only perform well but are good to society by promoting fairness and reducing inequality.

While most of my previous work has therefore established a very strong foundation of technical excellence, I feel that this scope actually needs to be broadened further by including ethical considerations and inequity. By embedding these principles at the core of my projects, I consequently believe that I help develop responsible, more equitable data-driven solutions.