Ethics in Data Collection

Course Overview

Summary

Before we even consider the use of the data, it's crucial to lay a solid ethical foundation in the very first step of the data journey: data collection.

In this course, we will explore three key areas and address the ethical nuances, potential sources of bias, and the importance of terminology in the data collection process. This will equip you with the knowledge and tools necessary to navigate the intricate terrain of data collection with ethics and integrity.

Learning Outcomes

Data ethics is considered across every stage of the data lifecycle. This course supports you in evaluating your own data collection processes to determine their effectiveness.

To achieve this, you will:

- Explore several aspects of data collection, such as types, sources and methods and investigate the potential impact of these approaches through an ethical lens
- Evaluate the sources of bias during data collection and mitigation strategies
- Look at the impact of bias in terminology and language and how data minimisation practices can play a key role in not only mitigating bias, but also in compliance.

Learning Experience

Number of modules	3 (+ reflective workbook)
Modality	Asynchronous / Self-directed / Online
Notional learning hours	2 hours (total)
Assessment	Formative
Certificate	Certificate of completion

Each module contains learning content that introduces the key concepts in the module, providing examples and case studies that demonstrate these concepts in practice. Each module contains a series of formative questions to support your learning. Learning is applied in activities throughout, and consolidates skills in the evaluation of bias during data collection and the potential mitigation strategies.

Module Summary

Module Name	Description
Ethical data collection: methods, sources and challenges	The initial step in any data-centric project involves data collection or acquisition. Whether the data already exists or is being gathered for the first time, the selection of methods and sources carries substantial implications for all subsequent project activities. Trade-offs are inevitable, often stemming from constraints such as time, budget, and project objectives. These trade-offs introduce a degree of risk. The pivotal question is, how many and to what extent are you willing to embrace these risks? In this section, we delve into various data collection methods and explore the challenges they present.

	- Primary and secondary data - Considerations for collecting data - Data re-use
Sources of bias during data collection	Bias can creep into the data collection process through various channels, from sampling techniques and terminology choices to the preference for quantitative or qualitative methods. Understanding these sources of bias is essential for ethical and accurate data practices. In this section, we delve into the intricacies of data collection and uncover the potential sources of bias that can affect the data we gather, regardless of its origin. In particular, this module looks att: - Defining bias in data collection - Understanding the impact of bias - Exploring where bias starts - Understanding the types of bias - Examining the potential for bias in data collection methods - Examining the potential for bias in sampling techniques - Examining the impact of targets - Evaluating and applying practices to mitigate bias
Navigating bias in terminology	Terminology is the foundation upon which data collection is built. The words and phrases used to define concepts, ask questions, and categorise responses have a profound impact on the data's quality, accuracy, and ethical integrity. In this section, we'll explore: - Terminology's influence on data and bias - How to identify biassed language - The strategies for selecting inclusive terminology - The role of data minimisation