

Name: _____ Date: _____

Raw Facts, Fine Print: Senior Data Ethics & Literacy Quiz

Can you spot the algorithmic bias in a dataset? Deconstruct complex data provenance and evaluate the socio-technical implications of information architecture.

1. A researcher examines a dataset of urban mobility patterns where data was only collected from users with high-end smartphones. This is an example of which data literacy concern?

- A. Survivorship Bias
- B. Sampling Bias
- C. Data Siloing
- D. Algorithmic Transparency

2. The concept of ___ refers to the chronological record of the origin, movement, and transformations of a dataset, essential for verifying its integrity.

- A. Data Normalization
- B. Data Scraping
- C. Data Provenance
- D. Data Warehousing

3. True or False: In a high-stakes predictive model, a high correlation coefficient (r) between two variables is sufficient evidence to establish a direct causal mechanism for policy-making.

- A. True
- B. False

4. When evaluating the 'Veracity' of Big Data in a corporate audit, which factor is most critical to investigate?

- A. The speed at which the data is processed
- B. The physical storage location of the servers
- C. The consistency and trustworthiness of the data points
- D. The file format of the raw metadata

5. To protect individual privacy in large public datasets, organizations often use ___, which adds 'mathematical noise' to the data to prevent de-identification.

- A. Differential Privacy
- B. Symmetric Encryption
- C. Data Sharding
- D. Boolean Filtering

6. Simpson's Paradox is a data phenomenon where a trend appears in several groups of data but ___ when these groups are combined.

- A. Remains identical

Name: _____ Date: _____

- B. Disappears or reverses
- C. Increases in statistical significance
- D. Becomes a linear regression

7. True or False: Using an 'unsupervised learning' algorithm for data analysis eliminates the risk of human bias being integrated into the final output.

- A. True
- B. False

8. An analyst uses a ___ to identify outliers in a dataset that might indicate sensor failure or fraudulent activity rather than genuine trends.

- A. Standard Deviation Test
- B. Data Lake
- C. Relational Schema
- D. Lookup Table

9. Which of these is a primary ethical implication of 'Data Persistence' in the context of the Internet of Things (IoT)?

- A. The difficulty of correcting inaccurate historical data
- B. The requirement for high-speed fiber optic cables
- C. The use of SQL over NoSQL databases
- D. The carbon footprint of physical data centers

10. True or False: Metadata (data about data) can often reveal more sensitive personal information in aggregate than the actual content of the primary data itself.

- A. True
- B. False