LARGE AMOUNTS OF DATA

BIG IDEAS:

- Organizations **collect data** to inform decisions and improve user experience
- Technology allows for the easy collection of **large amounts of data**
- Although there are numerous **benefits** to working with large amounts of data, there can also be several **challenges**

LEARNING GOALS AND SKILL DEVELOPMENT:
You know you have met the goals for this lesson when you can:

EMERGING

LEARNING GOALS	ANCHOR QUESTIONS
Give an example of a situation that would require collecting a large amount of data	1
List some benefits and challenges of using large amounts of data	2, 3
Describe some benefits and concerns of large amounts of user information collected by social media platforms	4

SKILL BUILDING QUESTIONS			
1	2	3	4

EVOLVING

LEARNING GOALS	ANCHOR QUESTIONS
Explain the meaning of biometric data and list some benefits and drawbacks of uses of biometric technology	5
Describe how data collection may be used in transportation	6
Give examples of how machine learning is being used in the real-world and list benefits and drawbacks of this technology	7

SKILL BUILDING QUESTIONS			
5	6	7	

EXTENDING

LEARNING GOALS	ANCHOR QUESTIONS
Explain the meaning of the three V's as they relate to big data and give examples of how big data is used in various industries	8

SKILL BUILDING QUESTIONS			
8			

BUILD YOUR SKILLS

1. Identify a current context that involves the collection and use of a large amount of data.

2. List some of the benefits of using large amounts of data.

3. List some of the challenges associated with the use of large amounts of data.

4. Many social media platforms collect data about their users' interests, behaviours and demographics (age, gender, location, etc.).

a) For what reasons might this information be collected?

b) Describe some potential concerns that users may have surrounding the collection and use of this data.

5. Biometrics are physical characteristics that can be used to identify individuals.

a) List some types of identification methods that involve biometric data.

b) List some possible uses of biometrics.

c) What are some advantages of using biometric technologies?

d) List some disadvantages or potential concerns associated with the use of biometric data.

6. Tesla, Inc. collects a variety of information from its customer's vehicles, including speed information, odometer readings, battery information, camera images, safety-related data, and other information from numerous sensors.

a) Suggest some ways in which the data collected could be used to benefit Tesla drivers.

b) Suggest some ways in which the data collected could be beneficial to the Tesla company.

c) Tesla uses the data collected from its vehicles to improve its self-driving system. List some potential challenges associated with the collection/use of data for this purpose.

7. *Machine learning* refers to a computer system's ability to learn for itself and improve itself based on experience and the analysis of data.

a) Give an example of how machine learning is utilized in the real world.

b) Predict some of the data-related challenges associated with machine learning.

c) Machine learning is a type of artificial intelligence. Describe some potential ethical issues surrounding the development and implementation of this technology.

8. The term *big data* is used to describe extremely large amounts of constantly growing data. Such data is too large or complex to be handled by traditional data-processing software.

a) Research and explain the meaning of the *three V's* (volume, velocity, and variety) associated with big data.

b) Determine one way in which big data is used in each of the following industries.

i) media and entertainment ii) healthcare iii) finance

ANSWERS

CHECK YOUR UNDERSTANDING

1. Answers may vary. For example, the censes survey conducted by the federal government, weather forecasting, video game user data used to improve user experience, and personal information collected by websites for marketing purposes.

2. Answers may vary. For example, larger amounts of data can increase the accuracy of representations and the reliability of predictions.

3. Answers may vary. For example, large amounts of data may be difficult to organize and manage. Large volumes of data require a great deal of storage space, especially if backups must also be stored. The processing of large amounts of data can require very high levels of computing power.

4. a) Answers may vary. For example, the data collected through social media platforms can be used to improve the overall user experience and to customize content/advertising for each user.

b) Answers may vary. For example, users might consider the collection of such data as an invasion of their privacy. They might also be concerned that their personal information will be misused or shared without their consent.

5. a) Answers may vary. For example, fingerprint scanning, facial recognition, voice recognition, and iris recognition.

b) Answers may vary. For example, forensic analysis, law enforcement, building access, mobile phone access and authentication, and airport security.

c) Answers may vary. For example, biometric data is difficult to steal or impersonate. It is also highly reliable and offers a convenient user experience.

d) Answers may vary. For example, individuals may have concerns about privacy and the safe storage of their biometric data. The use of biometric authentication is expensive. Other issues may include those related to accessibility, human rights, and discrimination.

6. a) Answers may vary. For example, the data can be used to efficiently diagnose or resolve issues with the vehicle (possibly before they happen), improve the vehicle's Autopilot system, analyze how well the vehicle is being maintained, and improve/customize the overall driver experience.

b) Answers may vary. For example, the data can be used to inform the company about how Tesla vehicles are being used. This information can help the company refine its systems and improve the performance and safety characteristics of future products.

c) Answers may vary. For example, privacy and security may be subjects of concern for drivers. The huge amount of data collected requires a massive amount of storage space. A very high degree of computational power is needed to process the data. Other concerns may be related to the overall safety of self-driving cars and their potential susceptibility to hijacking.

7. a) Answers may vary. For example, text-to-speech technology, email spam filters, and credit card fraud detection technology.

b) Answers may vary. For example, machine learning requires massive amounts of high-quality *training data*. Machine learning also demands a great deal of computing power and can take a long time before accurate results are delivered.

c) Answers may vary. For example, the implementation of machine learning may result in fewer job opportunities. Artificial intelligence may also be susceptible to bias and/or could potentially be used for malicious purposes. Another potential issue is the possible loss of human control over a complex intelligent system.

8. a) Volume refers to the huge amount of data. Velocity refers to the high speed at which the data is generated/collected and processed. Variety refers to the different sources and forms (structured, semi-structured, and unstructured) of the data.

b) i) Answers may vary. For example, streaming services use big data to provide recommendations based on users' viewing habits.

ii) Answers may vary. For example, big data is used to maintain patients' electronic health records and trigger important warnings/reminders.

iii) Answers may vary. For example, big data is used to predict financial trends and inform financial decision making.