



# Module 1: Understanding AI

## Summative assessment – Answers

**LO: Recognize that AI is a tool that can solve a range of complex problems**

**1. For each AI application below, which problem is solved using AI?**

**a. Driverless car**

		Feedback
Using sensors and a computer system to avoid pedestrians	✓	Yes! AI systems will use the input from sensors and a computer system to try to identify pedestrians, so the car can avoid them.
Making ethical decisions about what to do in different situations		No. AI systems cannot understand right and wrong, so they cannot make ethical decisions.



**b. Music streaming service**

		Feedback
Playing the music selected by the user on the app		No. This is a simple problem and can be solved using a rule written by a human.
Making recommendations based on a user's previous music selections	✓	Yes! This is a difficult problem, and the system needs data about the user's music choices to solve it.

**Misconception: Learners may have incorrect mental models around AI systems always being robots or generative AI systems, or around AI systems thinking or learning like a human.**

**LO: Use appropriate technical descriptions of AI systems**

**2. Which of the following options uses the correct language to describe AI?**

		Feedback
A. Social media apps use an AI to recommend content to users. This AI enjoys learning about you and understanding enough about you to decide what content you will like.		No. This description uses human behaviors like 'enjoy' and 'understand' to describe an AI application.
B. Social media apps use an AI model to recommend content to users. This AI model uses data about what you like to produce predictions about what content you will like.	✓	Yes. This description does not use human behaviors like 'enjoy' and 'understand' to describe an AI application.

**Misconception: Students may use human behaviors and characteristics to describe AI systems. This may lead to them developing an inaccurate mental model where they believe that AI systems think, feel, and learn in the same way that a human brain does.**



**LO: Describe the difference between 'data-driven' and 'rule-based' approaches to application development**

**3. What is the difference between a traditional computer program and a machine learning model?**

		Feedback
A. Machine learning models use rule-based IF/THEN statements.		No. Developers use data to create data-driven models when either the rules are not known or there are too many rules for a developer to create themselves.
B. Machine learning models use a data-driven approach.	✓	Yes. A machine learning model uses data-driven methods. Developers use a vast amount of relevant and correct data to train a model.
C. They are the same; the only difference is that machine learning models use more data.		No. Developers use data to create data-driven models when either the rules are not known or there are too many rules for a developer to create themselves.

**Misconception:** Students may believe that AI systems are created using the same rule-based approach that they may be using when writing algorithms in a procedural language like Python. They may not be aware that AI systems are created using a different paradigm that involves training a model on a large amount of data.

**LO: Describe the difference between 'data-driven' and 'rule-based' approaches to application development**

**4. Is this statement true or false?**

**A rule-based approach requires human developers and engineers, but a data-driven approach does not.**

		Feedback
True		No. Building a data-driven system requires a developer to make sure that the system is designed with the user in mind. To create a classification model, the developer will need to decide on the classes. They will also need



		to decide what data and how much data is used to train the model.
False	✓	Yes. Building a data-driven system requires a developer to make sure that the system is designed with the user in mind. To create a classification model, the developer will need to decide on the classes. They will also need to decide what data and how much data is used to train the model.

**Misconception: Students may not know the difference between a rule-based system and a data-driven system.**

**LO: Define the role of machine learning in creating models**

**5. How do developers use machine learning to create a data-driven model?**

		Feedback
A. By writing rules to solve problems based on patterns that are identified in large amounts of data		No. The problems that machine learning is used to solve could not be solved using a set of IF/THEN statements as they are too complicated. They require a developer to create a data-driven model.
B. By training a model to identify patterns in large amounts of data, which can then be tested and used to solve problems	✓	Yes! The problems that machine learning is used to solve could not be solved using a set of IF/THEN statements as they are too complicated. They require a developer to create a data-driven model.

**Misconception: Students may not be aware that a 'data-driven' system will use an algorithm to identify patterns in the data that the model is being trained on. Humans do not identify these patterns.**



**LO: Describe the impact of data on the output of a machine learning (ML) model**

**6. Some pictures produced by generative AI systems do not accurately represent the real world. Why is this?**

		Feedback
A. Generative AI is trained on patterns identified from many pictures. These patterns might be applied in unusual ways.	✓	Yes. When a generative AI application uses the patterns that it has identified from many pictures to generate new pictures, the patterns might be applied in unusual ways.
B. Generative AI is trained on patterns identified from many pictures. The only way to guarantee that pictures accurately represent the real world is to use more training data.		No. The pictures you train your model on will all have particular differences in terms of distance from the camera and position. These patterns are identified by the model and might be output in some unusual ways.

**Misconception: Students may believe that if the model is trained on a vast amount of data, the model will be accurate. It is important for them to know that even models trained on a huge amount of data may not be 100% accurate.**

**LO: Describe how supervised learning is used to create a classification model using classes and labeled data**

**7. When supervised learning is used to create a classification model, how is data prepared?**

		Feedback
A. Before the model is trained, the machine learning algorithm labels the data and puts it into classes.		No. The training data has to be labeled with its class by a human developer before the model is trained. During training, the model will identify patterns within the data in a class, to allow it to identify which class new data is likely to be in.
B. Before the model is trained, the human developer labels the data and puts it into classes.	✓	Yes! The training data has to be labeled with its class by a human developer before the model is trained. During training, the model will identify patterns within the data in a class, to allow it to



		identify which class new data is likely to be in.
C. Before the model is trained, the data is input into the model and the model then outputs labels for the data.		No. The training data has to be labeled with its class by a human developer before the model is trained. During training, the model will identify patterns within the data in a class, to allow it to identify which class new data is likely to be in.

**Misconception:** Students may believe that classes are automatically generated by the model. They may not realise that a human decides on the classes during the development of the model and then puts data into the classes. This training data is then used to create the model.

**LO:** Describe how supervised learning is used to create a classification model using classes and labeled data

8. The following diagram shows that a piece of data has been labeled as an orange. Which **one** of the following options best describes the term 'label'?

		Feedback
A. A category that data can be placed into		No. The category that the data is placed into is called the class.
B. Applied to a single piece of data to show which class it belongs to	✓	Yes. The developer gives each item of training data a label and this shows what class the data belongs to. Then, when the model is trained, it will identify patterns in the data in each class.
C. The confidence score that is returned by the model		No. The confidence score shows how likely it is that a predicted label is correct.

**Misconception:** Students may not understand the difference between labels and classes. A class is a collection of items. A label is applied to a single item to identify its class. In a supervised learning model, the model will produce a label for an item, along with a confidence score as a measure of probability that the label is correct.

**LO: Explain the need for both training and test data**

9. Which **one** of the following options best describes what test data is?

		Feedback
A. Data that is removed from the training data and used to check the accuracy of the model	✓	Yes. The test data comes from the same set as the training data but is not used to train the model. The developer knows which class the data should be labeled as, so the data can be used to test the confidence scores output by the model.
B. Data that is used to train the model		No. The test data comes from the same set as the training data but is not used to train the model. The developer knows which class the data should be labeled as, so the data can be used to test the confidence scores output by the model.
C. New data, which is not from the same set as the training data, which is entered into an application that uses the model, to check the accuracy of the model		No. The test data comes from the same set as the training data but is not used to train the model. The developer knows which class the data should be labeled as, so the data can be used to test the confidence scores output by the model.

**Misconception: Students may not realize that test data has been removed from the original training data, and that as a result, the developer will know which label the model should output for this data during testing. They may think that the test data was included in the training data. This would lead to overestimating the model's accuracy.**

10. A machine learning model has been developed to identify African birds from sound alone.

**LO: Describe the impact of data on the accuracy of a machine learning (ML) model**



**Part A:**

When developers tested the model, they found that many confidence scores were very low. Which **two** options might be the reasons for this?

		Feedback
A. They may not have used enough sound recordings of different types of African birds in the training data.	✓	Yes. If the machine learning model has not been trained on enough data, the outputs may often be incorrect.
B. Machine learning models cannot be trained very well with sound data.		No. Machine learning models can be trained with lots of different types of data, for example, images, text, numbers, or sound.
C. The sound quality of the recordings may not have been good enough.	✓	Yes. If the sound recordings were low quality or there was too much background noise in them, this could affect the model's accuracy.

**Misconception:** Students may think that 'data-driven' systems can only be trained effectively with text, numerical, or image data.

**LO:** Explain how bias can influence the predictions generated by an ML model

**Part B:**

While testing and evaluating the model, developers have found that the model cannot identify a particular type of bird. What should the developers do to improve their model? Select **one** option.

		Feedback
A. Add more sound recordings of all known types of birds		No. Adding more sound recordings of all known types of bird will not help. The training data for the model needs to include more sound data from that particular type of bird. Adding more sound recordings of that type of bird will allow the model to identify data patterns in that type of bird's call.



<p>B. Add sound recordings of humans and other animals in different environments</p>		<p>No. Adding more sound recordings of humans and other animals will not allow the model to identify data patterns in the call of one particular type of bird.</p> <p>The training data for the model needs to include more sound data from that particular type of bird. Adding more sound recordings of that type of bird will allow the model to identify data patterns in that type of bird's call.</p>
<p>C. Add sound recordings of this particular bird in different conditions, such as close up and far away</p>	<p>✓</p>	<p>Yes. The training data for the model needs to include more sound data from that particular type of bird. Adding more sound recordings of that type of bird will allow the model to identify data patterns in that type of bird's call.</p>

**Misconception: Students may believe that the only way to improve a 'data-driven' system is to increase the amount of data it is trained on, and not realize that they also have to consider the type and quality of that data.**

