6.DA.4 Data and Analysis

> *The student will identify ways people curate and provide training data. (a) Identify and list ways people provide data that is used as training data. (b) Discuss the role of human intervention in curating training data. (c) Identify and describe the effect training data has on the accuracy of artificial intelligence systems.*

Understanding the Standard

Artificial intelligence technologies, these days, use a process called "machine learning" to analyze data and make predictions, with the ultimate goal of using patterns in the data to imitate human work. The data used as the foundation of these predictions is called "training data". Companies making machine learning technologies can collect this data in all sorts of ways, from scraping data available on the public Internet to collecting data from individual users like on social media. The training data has a big impact on the predictions the machine learning technology can make, and many times bias present in the training data results in biased predictions (see VDOE CS standard 6.CSY.3 for a specific example of how bad training data can lead to poor outcomes, including "algorithmic bias"). In order to create machine learning models that perform well, humans need to curate training data to ensure that the model ends up identifying useful patterns and correlations in the data.

This standard is very closely related to 6.CSY.3, which covers almost exactly the same concepts explicitly in the context of artificial intelligence.

Term	Definition
Machine learning	The process by which a computing technology ingests data, analyzes it, and makes predictions.
Training data	The set or collection of data used to train an AI model to recognize patterns or make decisions.

Prerequisite Knowledge

Students should have a basic understanding of "bias", "AI", and "machine learning" in broad terms before engaging with this standard.

Integration Opportunities

English 6.C.4b.c Deconstruct media messages using elements of media literacy to explore how data can be collected and used to create more targeted, intentional, and effective messaging.

Physical Education 6.3b,e,f Identify various ways fitness can be measured with or without technology. Discuss the importance of varying fitness measures and use information to help create a personal activity plan.

Math 6.PS.1 Research different sources of training data for AI, create circle graphs to represent the proportions of each source, and analyze the potential algorithmic biases and impacts on AI accuracy, discussing the importance of human intervention in curating data.

Summary of a Lesson

Have students list technologies they think might be "artificial intelligence". Then, introduce the main characteristics of AI: collecting data, analyzing data, and making predictions. Have students edit their list to add other technologies that might meet these criteria and remove technologies that don't. Facilitate a discussion where students identify what data the AI technologies use to make predictions. Then, brainstorm ways that data might contain bias that would result in bad predictions. You can address this standard in a hands-on way by having students create simple machine learning models using tools like Teachable Machine, which do not require any coding. Check out this lesson plan for an example of how you might facilitate a lesson like this.