Appendix A
Sample AI Literacy Curriculum 3

Today's mission
You will be the key stakeholders to design solutions to minimize the real-world algorithmic bias problems in each scenario.

- Scenario 1: Your family want to categorize waste, but you always feel distressed to distinguish the category of waste. You want to develop an AI that helps in categorizing waste into different categories like organic, recyclable, and non-recyclable.
- Scenario 2: Your parents want to develop an AI system that can scan and determine whether a product (e.g., designer handbags, sneakers) is authentic or counterfeit by examining its features, tags, and materials.
- Scenario 3: A scientist want to develop an AI tool that identifies and classifies various urban wildlife species from CCTVs to assess biodiversity, track species movements, and aid in urban planning

Guiding questions:

- What classes would you create?
- What features might characterize each class?
- Where would you get your data for the classes?
- What data would you need to include to make sure it is representative of the diversity in the class? 5
- What possible algorithmic biases could occur in this classifier?
- How would you test your classifier to make sure it isn't biased?
- What might be the negative consequence if this classifier is biased

EXAMPLE

Scenario Example
You want to design a spam filter that will recognize spam text and automatically move them to the trash folder.

What classes would you create?
Class 1: Spam
Class 2: Not spam

What features might characterize each class?
SPAM:

- Sent to a list
- Sales language like "buy now" or "earn money"
- Overuse of exclamation points
- Overuse of capital letters
- Asking for money
- Asking to forward the message
- Sending from an unknown number
- Including a suspicious link

NOT SPAM

- Sent to an individual

- Recipient's phone number is recognized
- Sent from someone in the user's contacts
- From a subscription service the user has agreed to
- Proper spelling and grammar

Where would you get your data for the classes?
Think about where you might find examples of text messages that are labeled as spam or not spam.
One way to collect data would be to have a large group of users mark spam text messages they receive as spam and leave non-spam messages alone, then use these as the two training data sets.

What data would you need to include to make sure it is representative of the diversity in the class?
Think about different types of text users (age, gender, interests) and the types of spam and non-spam texts they might receive.

The spam training data would need to include many types of spam texts — ones that ask for money, ones that are a sales hoax, ones that ask the user to forward it, etc.

The non-spam training data would need to include examples of texts that are received by all types of message users, including those from:

- Children, students, and adults
- Both professional and friendly messages
- Legitimate messages about sales or banking

What possible algorithmic biases could occur in this classifier?
Think about ways that the classifier might work unfairly.
If there aren't enough diverse examples of non-spam, the spam filter might filter out all messages that have spelling or grammar mistakes, even from friends, or it might filter out all sales or money-related messages, even from places the user shops or banks at.

How would you test your classifier to make sure it isn't biased?
Think about how to make sure you have quality test data.
You might test the classifier with a large dataset that has diverse data for both classes that represents all the possible types of text messages that someone might receive.
You should test your classifier to make sure it consistently classifies spam and non-spam text messages correctly.

What might be the negative consequence if this classifier is biased?
If spam texts get through the filter, then the user might receive messages that include phishing links, malware, fake sales, or other hoaxes.
If the filter labels too much as spam, then the user would miss important messages from family, friends, work, or businesses.