AI for Children

Artificial Intelligence Curriculum for Elementary and Secondary Schools

Card

Bias

03 Machine Learning Deck

https://kurikulum.aidetem.cz/cards

Created by: Radek Špáta & Eva Nečasová
Expert guarantors: Tomáš Mlynář, Pavel Kordík
Methodological consultants: Peťa Dovhunová, Ondra Brém
Language proofreading: not yet done
Last update: 01/2025
Version: 04

These teaching materials were translated using ChatGPT.
Please note possible imperfections in the expressions or wording.

Form for comments

Teaching Material for the AI Curriculum for Elementary and Secondary Schools
Computer Science at Secondary Schools - Cards, Machine Learning Deck

Bias

Basic information about bias

Imagine you're building an AI app that can recognize shoes in pictures. First, you'd need to show it a large number of shoe images so it can learn what shoes typically look like—laces, soles, and so on. Now try to picture as many different types of shoes as you can. You'd show all of those to the AI. But here's the catch—did you imagine not just all the different types of shoes, but also all the possible angles, colors, materials, and tiny details? What about shoes worn in ancient Egypt or 5th-century China? AI doesn't see the world the way we do. It can't touch objects or fill in the gaps the way people naturally can. It struggles with generalizing from limited examples. So if you forget to include certain types of shoes in its training, it might not recognize them as shoes at all. That's what we call bias.

Expanded definition for advanced users

Artificial intelligence systems learn from data prepared by humans. If the data is poorly selected, unbalanced, or incomplete, the AI can become biased or distorted in its decision-making. This is known as bias. Bias can take many forms. One example is historical bias—longstanding human prejudices, like those based on race or gender, may be reflected in the data. For instance, if men have historically held more leadership positions, the AI might learn to recommend men for those roles more often. Another form is representational bias. Imagine a facial recognition system trained mostly on faces from Central Europe—will it work just as well in Africa? Probably not, because it hasn't learned to recognize a wide enough variety of faces.

If AI is going to play a larger role in decision-making, we must be able to trust it to be accurate and fair. That's why it's crucial to train these systems on the right kind of data.

Machine Learning Deck

Dataset
Supervised Learning
Bias
Reinforcement Learning

Lesson presentation in PDF

Editable presentation in Canva

Note: Gender equality is key for AI for Children, but for brevity we use masculine formulations in our methodologies.

Classroom Activity

Alien Detective Agency – Part III

45 min

Activity description

In the previous exercise, we trained a machine learning model using Teachable Machine with images of two alien families–Fluffs and Earls. We expected the model to classify new-found aliens with high accuracy by assigning them to the correct family. However, we discovered that the model relied heavily on one specific feature–ears. That's why it classified our first alien as a member of the Earl family. We were able to test this theory by trying the model on images of other aliens. An alien without ears was classified as a Fluff. In other words, even though our first alien shared many attributes with the earless one (and seemed to belong to the same family), the model ended up assigning them to different groups.

This activity continues directly from the previous task in the "Supervised Learning" lesson. In this hands-on exercise, you'll use a pre-prepared project in Teachable Machine. Thanks to a new expedition on the alien planet, two more families will be discovered and added to the training data. However, even with the expanded dataset, the model will still misclassify the aliens–its bias remains. Only a bold journey into the forbidden regions of planet Yemi will finally lead to change.

How the detective agency relates to bias

When we train a machine learning model on data that's unclear or too limited–like the images from just two alien families–it becomes very difficult for the model to correctly classify new data, such as unidentified aliens. Even when we expand the dataset by adding two more families, the problem doesn't go away. The model still relies heavily on one specific feature–ears–to make its decisions. As a result, it continues assigning aliens to the wrong families. But in the "real world," ears might not actually be the right feature to focus on. It could be fur, or some other detail we didn't even notice. And yet, the model makes its decisions based on that feature. That's what makes it biased.

Lesson Overview

Recommended Age, Lesson Length
Children aged 11–15, 45 minutes.

Building Blocks
Bias, dataset, machine learning.

What Are the Students Learning?
Machine learning systems can be biased. Bias can be reduced through model testing and dataset refinement.

Why Are They Learning This?
The goal of this lesson from the Machine Learning card set is:
To critically evaluate decisions made by artificial intelligence systems.

How Do We Know They Have Learned It?
They will explain the concept of bias.

Tools
Teacher: Sets of alien family cards, projection equipment, presentation.
Students: Computer, laptop or tablet (OS doesn't matter) for each student or group.

Digital Competence
Communication and Collaboration

Bloom's Taxonomy
Understanding: Understand how insufficient or poorly representative data can lead to bias in AI models.
Application: Use the Teachable Machine tool to train and test a model, identifying cases of bias in the data.

Five Big Ideas
3-C-III Datasets (Bias).

Engage

10 min

Recall

In the previous lesson, we trained a machine learning model in Teachable Machine using data from two alien families–Fluffs and Earls. We expected the model to correctly assign each foundling to the right family. However, we discovered that ears were the key distinguishing feature the model relied on. How did we find that out?
We tested our theory by running the model on images of other foundlings. The earless foundling was the only one classified as a Fluff. In other words, even though our first foundling shared many features with the earless one (suggesting they might belong to the same family), the model still classified them as members of different groups.

Explain

We trained our model in Teachable Machine using a method called supervised machine learning. But how does this type of learning actually work?
Machines, much like humans, can learn from examples. One of the ways they do this is through a method known as supervised learning. Simply put, it means we don't need to give the machine exact instructions or step-by-step procedures (an algorithm). Instead, we provide it with examples, and it learns how to perform certain tasks on its own. In this process, humans act as teachers–we tell the model what each example represents, often by labeling the data or sorting it into categories.

Think about it.

We still don't know which family the unidentified alien belongs to. So what could help us?
We could expand the dataset by adding more images of aliens. It would also help to have more detailed information about both the unidentified alien and the existing families. Another option is to classify the alien using different types of data–not just images, but for example DNA or other biological information.

Understand

25 min

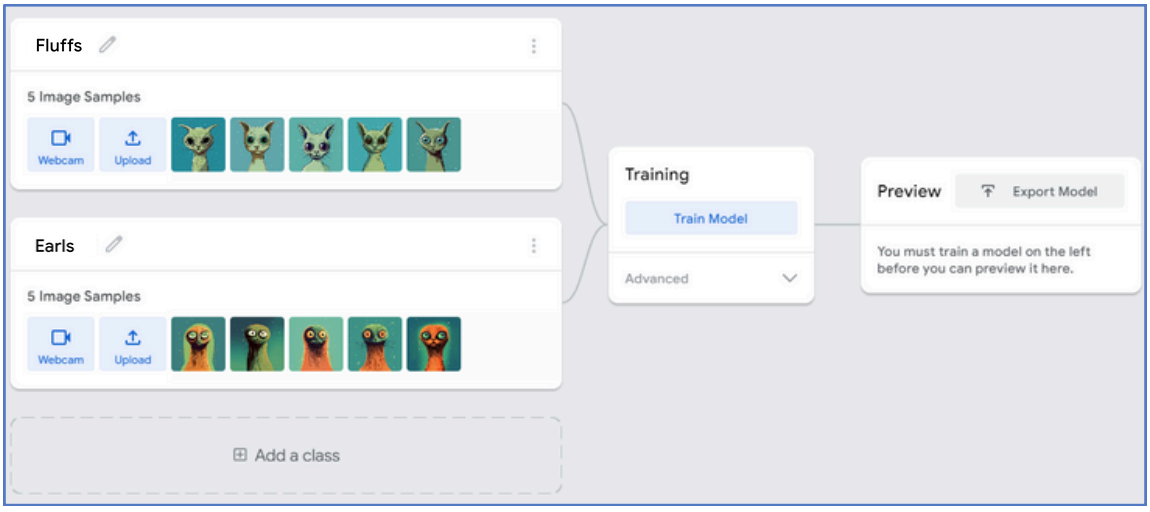
5 min

Activity 1

2-families.tm

Presentation slide 02

Students open the Teachable Machine website: teachablemachine.withgoogle.com
→ Select Get started → Open an existing project from a file.
→ Share the pre-prepared project file "2-families.tm" with them – bit.ly/aliens_03.
→ Students can also scan the QR code in the presentation on slide 02.
→ Once uploaded, they'll see a project with two classes – Fluffs and Earls.



Presentation slide 03

The story continues... (in the presentation on slide 03).
Local media are reporting a wave of new findings from the expedition into the unexplored regions of planet Yemi. It appears that humans may have discovered entirely new alien families! One of our colleagues at the detective agency is also working with the expedition team. Thanks to this connection, we now have exclusive access to brand new evidence!

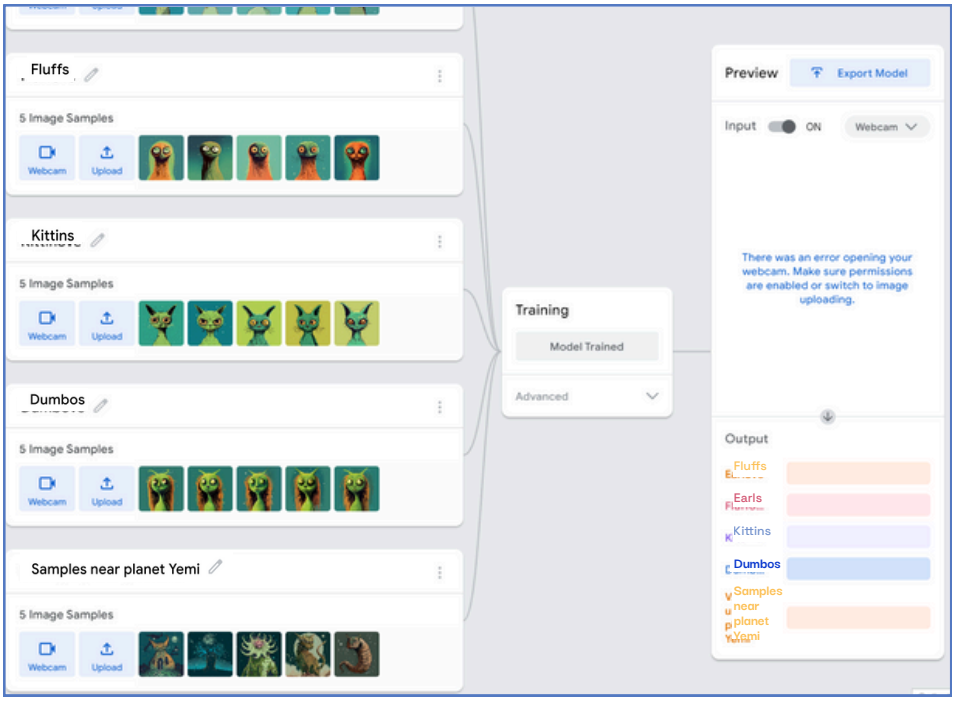
10 min

Presentation slides 04 and 05

Share the new data with students (link and QR code in the presentation on slide 04).
Guide students to explore the new data provided by the expedition. Can any of these samples help improve our model and solve the mystery of which family the unidentified alien belongs to?
The data can be found at: bit.ly/expedition-findings

new data

Students will add to the dataset in Teachable Machine. They should create new classes, name them, and upload the corresponding images. As they work, pay attention to whether they're selecting appropriate images (i.e. aliens), or if they're mistakenly including unrelated pictures such as plants or other animals. Let students create their dataset however they choose. After they train the model and try classifying the foundling, you can work with any misclassifications to analyze why the neural network still assigns the alien to the "wrong" class. This could include improving the new class or removing unsuitable images from it.



5 min

Presentation slide 06

Students test the trained model using the original alien images
Students test their model using the original alien images from the previous lesson (bit.ly/foundlings). Most models now classify the aliens as part of the Kittin family. And what happens if they show their own face to the camera? 😉

foundlings

Presentation slide 07

Show students the explanation of bias on slide 07 of the presentation.
Talk with students about when machine learning models can be biased. Can they think of examples similar to the one with shoes? Examples might include fruit vs. vegetables, European vs. Asian faces, or any situation where the model is trained on limited or unbalanced data.

Reflect

10 min

Share & argue

Which class did your model assign the foundling to, and why?
The answer is likely the Kittin family. The earless foundling was still being classified as a Fluff.

Do you think the ear feature played a key role again?
Based on our testing, even with the new data, ears remained the deciding factor.

The model worked technically (it tried to recognize images), but we still can't say whether it assigned the foundling correctly. How is that possible?
We don't have enough data and we don't actually know the ground truth. That means we can't confirm whether the model's classification was right or wrong.

Based on your experience training and testing a machine learning model – do you think we can fully trust AI systems? Why or why not?
AI builds its understanding based on the data we give it. If that data is flawed or not thoroughly tested, the model might not work correctly. That's why it's so important to think critically about decisions made by AI systems.

How can we tell if an AI system (model) is biased?
Through careful testing.

So what was the truth about the foundling? Which family did it "really" belong to?
Story ending: During the expedition, the alien's real family hadn't been discovered yet, which means there was no data about them in the dataset. That's why the model kept assigning the alien to the wrong families. Later, the foundling's actual family came forward–after they received a signal calling all families on planet Yemi to check in. So in the end, the foundling found their real home!

Presentation slide 08

Show students slide 08 in the presentation.🥰

Follow-up lessons

In this lesson, students learned what bias means in the context of artificial intelligence.
The story of the Alien Detective Agency now comes to an end. The final lesson in the Machine Learning unit will follow, introducing students to another type of machine learning – reinforcement learning.
Unlike supervised learning, where machines learn from examples, reinforcement learning teaches machines through experience – by trying things out, learning from mistakes, and improving over time.