

Train the *Machine*

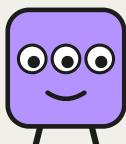
Students become a machine-learning model. They learn a rule from labeled examples, fail a test, and discover the real lesson: AI is only as good as its data. That's also how kids can help make AI better.

GRADES 3-10

45-55 MINUTES

NO DEVICES

AI LITERACY • SCIENCE • DATA



WHICH ONES ARE ZIBS? YOUR TRAINING DATA DECIDES.

WHAT'S INSIDE

- Teacher guide with scripts
- 9 test cards
- Student model notebook
- "How kids improve AI" poster page
- 12 training cards, set A (skewed data)
- 12 training cards, set B (diverse data)
- Real-world reading: shortcuts & bias
- Answer key, standards & sources

THE BIG IDEA

An AI learns whatever its *data* teaches it

Most AI isn't programmed with rules. It's **trained**: it sees thousands of labeled examples ("cat," "not a cat") and finds a pattern that separates them. If the examples are lopsided, the AI can learn the wrong pattern, a **shortcut**, and still look smart until it meets something new. In this lab, students are the model.

THE SECRET (TEACHER ONLY)

A **Zib** is any creature with **exactly three eyes**. Color, shape and antennae don't matter.

In **training set A**, every Zib happens to be blue and every non-Zib is orange or green. Most students learn "blue = Zib." That rule scores only **3 out of 9** on the test.

Training set B adds diverse examples (Zibs in every color, blue non-Zibs), and students discover the real rule.

HOW THE LAB MAPS TO REAL AI

Labeled cards	training data
Your rule	the model
Finding a rule	training
Unlabeled cards	test data
Score out of 9	accuracy
"Blue = Zib"	a shortcut / bias
Set B	better, more diverse data

Students will be able to

- Describe how a model learns from labeled examples.
- Calculate accuracy on test data.
- Explain how skewed data creates biased or wrong AI.
- Name ways people, including kids, make AI better.

Materials & prep (15 min)

- One set of cards per team of 3–4 (set A, test, set B), each cut out and kept in separate envelopes labeled A, TEST and B.
- One model notebook per student.
- Optional: print in color. The lab still works in grayscale because students can write the color on each card.

- 1 **Train (8 min).** Teams open envelope A. *“These are labeled examples. Your job is to be the AI. Find a rule that tells a Zib from a not-Zib, and write it in your notebook as Rule v1.”*
- 2 **Test (7 min).** Teams open the TEST envelope and use *only* their rule to predict each card. No changing the rule now! Then read the answers aloud (key on the last page) and teams calculate accuracy: correct ÷ 9.

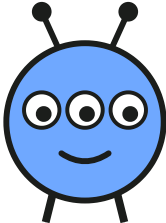
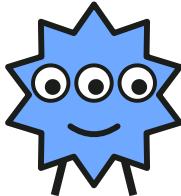
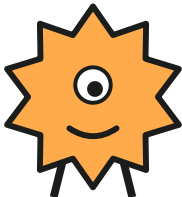
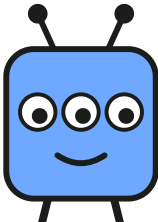
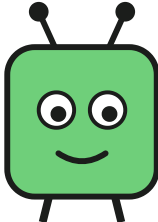
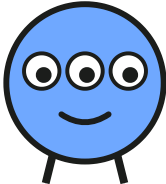
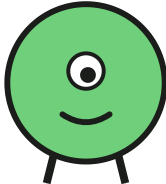
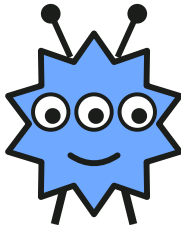
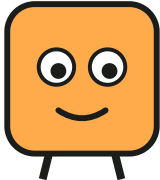
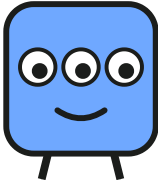
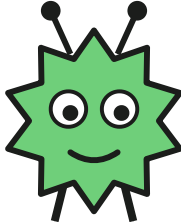
“Most of you scored 3 out of 9. That’s worse than flipping a coin! But your rule was perfect on the training cards. What happened?”
- 3 **Diagnose (7 min).** Teams list the test cards they got wrong and look for what those cards have in common. Lead them to the problem: *every Zib in our training data was blue, so the model learned color.* Ask: was the model wrong, or was the **data** wrong?
- 4 **Retrain (8 min).** Teams open envelope B, combine it with A, and write Rule v2. If they’re stuck after 4 minutes, give the hint: *“Count something.”*
- 5 **Retest (5 min).** Teams predict the test cards again with Rule v2 and calculate the new accuracy (it should be 9 out of 9).
- 6 **Connect (10 min).** Read the real-world page together, then go through “How kids make AI better.”

“The AI didn’t get smarter by itself. It got better because people gave it better examples. Collecting, labeling, testing and reporting are jobs people do, and you can do them too.”

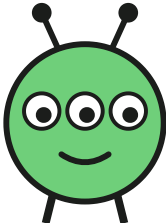
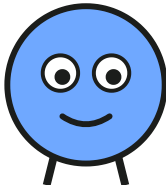
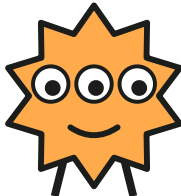
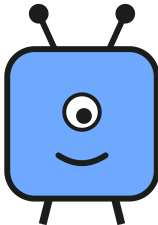
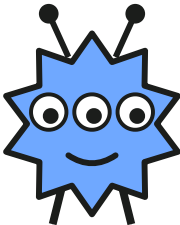
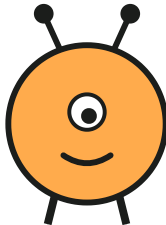
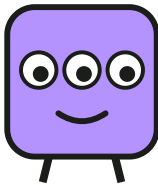
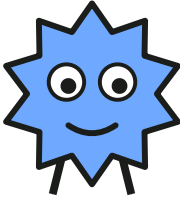
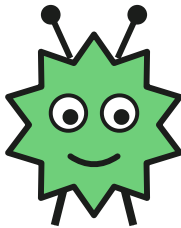
Adjust it

Grades 3–4	Do it as a whole class with the cards on a document camera. Skip fractions and just count how many it got right.
Grades 5–8	Run it as written. Turn accuracy into a percentage ($3/9 \approx 33\%$, $9/9 = 100\%$).
Grades 9–10	Have teams draw a tricky new test card (one eye hidden behind a leaf, say) and swap it with another team. Discuss “label noise” and who decides the labels.

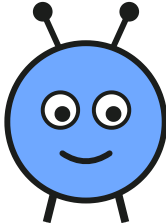
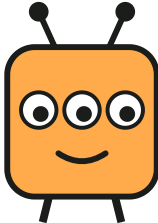
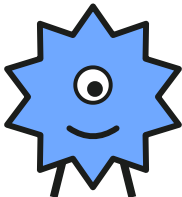
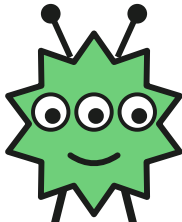
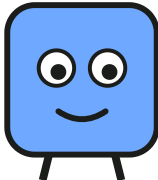
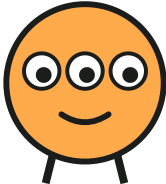
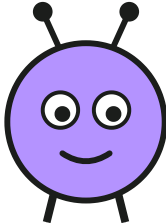
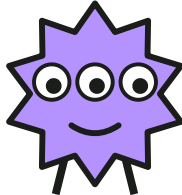
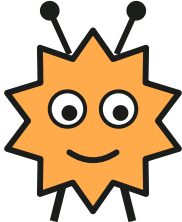
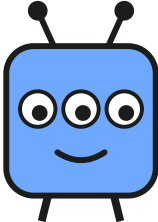
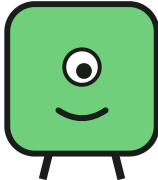
Envelope A. Twelve labeled examples. Cut along the dashed lines.

× A1 ✓ ZIB 	× A2 × NOT A ZIB 	× A3 ✓ ZIB 
× A4 × NOT A ZIB 	× A5 ✓ ZIB 	× A6 × NOT A ZIB 
× A7 ✓ ZIB 	× A8 × NOT A ZIB 	× A9 ✓ ZIB 
× A10 × NOT A ZIB 	× A11 ✓ ZIB 	× A12 × NOT A ZIB 

Envelope TEST. Predict each card with your rule. Don't change your rule while testing!

× T1 ? ZIB OR NOT 	× T2 ? ZIB OR NOT 	× T3 ? ZIB OR NOT 
× T4 ? ZIB OR NOT 	× T5 ? ZIB OR NOT 	× T6 ? ZIB OR NOT 
× T7 ? ZIB OR NOT 	× T8 ? ZIB OR NOT 	× T9 ? ZIB OR NOT 

Envelope B. Twelve more labeled examples. Combine them with set A to retrain your model.

× B1 ✓ ZIB 	× B2 × NOT A ZIB 	× B3 ✓ ZIB 
× B4 × NOT A ZIB 	× B5 ✓ ZIB 	× B6 × NOT A ZIB 
× B7 ✓ ZIB 	× B8 × NOT A ZIB 	× B9 ✓ ZIB 
× B10 × NOT A ZIB 	× B11 ✓ ZIB 	× B12 × NOT A ZIB 

Name: _____

Team: _____

Date: _____

My model *notebook*

1 · Train: Rule v1

A creature is a Zib if...

2 · Test with Rule v1

Write Z (Zib) or N (not a Zib) for each test card. Then mark each one ✓ or ✗ when your teacher reads the answers.

CARD	T1	T2	T3	T4	T5	T6	T7	T8	T9
Prediction									
✓ / ✗									

Accuracy v1 = correct ÷ 9 = ____ ÷ 9 = ____ %

3 · Diagnose

What do the cards you got wrong have in common? Why did your model learn the wrong rule?

4 · Retrain: Rule v2

After seeing set B, a creature is a Zib if...

5 · Retest with Rule v2

CARD	T1	T2	T3	T4	T5	T6	T7	T8	T9
Prediction									
✓ / ✗									

Accuracy v2 = ____ ÷ 9 = ____ %

6 · Reflect

What made the biggest difference: a smarter model, or better data? Explain.

REAL WORLD

This happens to *real* AI

WOLF OR HUSKY?

In a well-known 2016 experiment, researchers trained a model to tell wolves from husky dogs. It seemed accurate, but when they checked what it was looking at, it had mostly learned **“snow in the background = wolf.”** The training photos of wolves happened to have snow. It was a shortcut, just like “blue = Zib.”

Ribeiro, Singh & Guestrin (2016), “Why Should I Trust You?”, KDD.

SKIN-CHECK APPS

AI can help doctors spot skin cancer in photos. But dermatologists warned that many of the image collections used for training contained mostly lighter skin, so tools might work less well for patients with darker skin. The fix is more diverse, carefully labeled data.

Adamson & Smith (2018), *JAMA Dermatology* 154(11).

How kids make AI *better*

1 · *Collect*

Gather examples from lots of different people, places and situations, so the AI doesn't learn a shortcut.

2 · *Label*

Label examples carefully and honestly. One wrong label teaches the AI something false.

3 · *Test*

Try the AI on new, unusual cases and find where it breaks. A failure you find is a gift.

4 · *Report*

When an AI is wrong or unfair, say so: use the feedback button, tell an adult, or write to the company.

Real projects kids can join: On **iNaturalist**, people photograph plants and animals, and those community-identified observations help train its species-identification AI. On **Zooniverse** projects like Galaxy Zoo, volunteers label galaxies and other science images that researchers use, including to train AI. (Check each site's age rules. Younger students can take part through a class or teacher account.)

Discuss

- Who decides what the “right” label is? What if people disagree?
- Think of an AI you use. Whose data might be missing from its training?

Answer *key*

CARD	T1	T2	T3	T4	T5	T6	T7	T8	T9
Correct	Z	N	Z	N	Z	N	Z	N	N
"Blue = Zib"	N	Z	N	Z	Z	N	N	Z	N

The true rule is **exactly three eyes**. The color rule gets T5, T6 and T9 right, so it scores **3/9 ≈ 33%**. The eye rule scores **9/9 = 100%**. Card T7 is **purple**, a color the model never saw in set A. That's a great prompt: *"What does an AI do with something it has never seen before?"*

OTHER RULES STUDENTS MAY TRY

- **"Blue and has antennae"**: fails even on set A, so ask them to check their rule against all the training cards first.
- **"Not orange or green"**: works on A but fails T7 and gets the blue non-Zibs wrong. Same shortcut, said differently.
- **"3 eyes"**: some students notice it in set A. Celebrate it, and ask how they could be *sure* with so little evidence.

Standards alignment

FRAMEWORK	CONNECTION
AI4K12 Five Big Ideas	#3 Learning (models learn from data) and #5 Societal Impact (bias, fairness)
NGSS Science & Engineering Practices	Analyzing and interpreting data; constructing explanations; engaging in argument from evidence
CCSS Math 6.RP / 7.RP	Ratios and percentages (accuracy)
CSTA Data & Analysis strand	Using data to make predictions, and how data collection affects results

Sources

- Ribeiro, M. T., Singh, S. & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proc. ACM KDD*.
- Geirhos, R. et al. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence* 2, 665–673.
- Adamson, A. S. & Smith, A. (2018). Machine Learning and Health Care Disparities in Dermatology. *JAMA Dermatology* 154(11), 1247–1248.
- Interactive version: arqiv.si/teachers/labs

Thank you for teaching with ARQIV

This resource was made to help students think first and then learn to use AI well. We hope it sparks great conversations in your classroom.

You may

- Print and copy this resource for the students in your own classes.
- Display it on your classroom projector or board.
- Share it with students through a password-protected learning platform (Google Classroom, Canvas, Schoology and similar).

Please don't

- Share it with other teachers, a whole school or a district. Send them to **arqiv.si/teachers/resources** instead, or ask us about a school license.
- Post it on a public website, or sell or redistribute it in any form.
- Use it to train or fine-tune an AI model.

Questions, school licenses & workshops

Email **steve.ryan@arqiv.fm**. We also run Think-First workshops for staff and families.

FREE ONLINE COMPANIONS

Interactive versions of our games and labs are at **arqiv.si/teachers/labs**. Students can use them on any device, with no login.

ARQIV resources are written by people and reviewed for accuracy. Facts were checked against the sources listed in each kit as of 2026. AI changes quickly. If you spot anything that needs updating, please tell us.