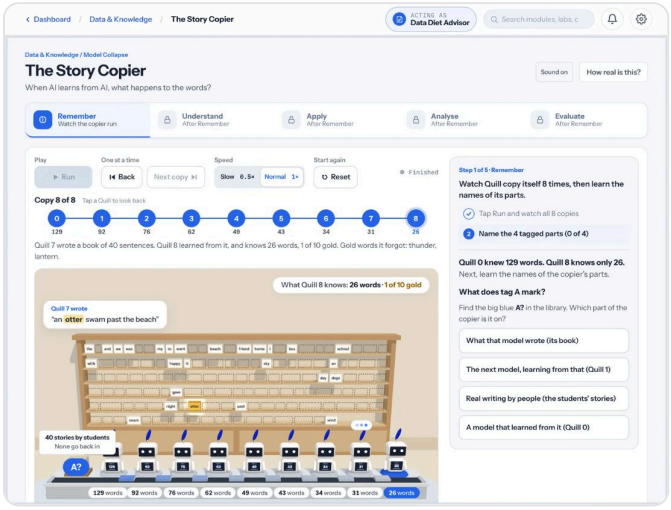
Module 04 · Data & Knowledge · Unit 04 · Stage 1, Foundations

The Story Copier

When AI Learns from AI

Grades 7–10 | **25 minutes** | Role: **Data Diet Advisor**

Grade 7: About age 12–13 · Year 8 in the UK · Class 7 in India



The big idea

When an AI learns from text another AI wrote, it loses the rare things first, and once they are gone they never come back. Quill forgets rare words like "lighthouse" generation by generation. Bigger books slow the loss, a tidy-up filter speeds it up, and putting the students' own stories back in each round keeps the rare words.

The learner's job

Learners are the **Data Diet Advisor** at Larkfield School · Story Bot project. Students taught a robot to write. Next year's robot will learn from this year's robot's books. Their task: decide what the school's story robot learns from, so it never forgets its rare words.

By the end, learners can say

- I can name the loop: human stories, a model, its book, the next model
- I can predict which words survive copying: rare ones vanish first, common ones last
- I can choose a training recipe that keeps rare words for eight generations
- I can compare two copiers and name the setting that made one forget more
- I can judge whether a real kind of app will lose rare things, backed by evidence from your runs

Before class

- One computer or tablet for each learner or pair, with EdSutra open (this Module is part of the full EdSutra studio).
- Open the studio once on each device before class, so the Unit still runs if the internet drops.
- Play the Remember step yourself, so you know where learners will click and what the four tags mark.
- The copier plays sounds. Bring headphones, or use the Sound button at the top of the Unit to turn them off.
- Best after How Machines Learn and Data & Datasets and Human-in-the-Loop.
- Units open in order, and Module 1 comes first. To start a class here, switch on Open every Unit in Settings → Teachers & parents.

Words to introduce

Training data: The text a model learns from.

Synthetic data: Text or pictures that an AI made, not a person.

Generation: One round of copying: Quill 3 learned from Quill 2's book.

Model collapse: Each generation knows fewer words; rare ones go first and never return.

Rare (gold) word: A word the students used only once, shown in gold.

Tidy-up filter: A setting that drops sentences with unusual words before the next Quill learns.

Running the lesson 29 minutes

Open 4 min

"Students taught a robot to write. Next year's robot will learn only from this year's robot's books. What might it forget?"

Before anyone clicks, ask: after eight robot copies, which words disappear first, common ones like "the" or rare ones like "lighthouse"? Take a vote and return to it in Understand.

Step	What learners do	Ask the class	Listen for
1. Remember 4 min	Learners press Run and watch Quill copy its book eight times while the word counts fall. Then they name the four tagged parts: real writing, the model, its book and the next model.	• Where did Quill 0 learn to write? • What does Quill 1 learn from, and what does it never see?	Quill 0 learned from the students' stories. Every Quill after it learns only from the last Quill's book.
2. Understand 4 min	Three rounds: learners guess whether a Quill still knows a word, then run to check. "Lighthouse" by Quill 3, "the" by Quill 8, then "lighthouse" with all 40 stories added back.	• Why did "lighthouse" vanish so early, while "the" lasted? • What changed when every Quill also read the students' stories? • Look back at our vote. Were we right?	A word used once can miss one book and never return. Common words appear in every book, so they last.
3. Apply 6 min	The control desk opens. Learners change the recipe (students' stories each round, sentences per book, the tidy-up filter) and run until at least 8 of the 10 gold words last all 8 generations.	• What is the smallest share of students' stories that kept the gold words? • Did bigger books alone save them?	Only putting enough real stories back in each round keeps the rare words. Bigger books just slow the loss.
4. Analyze 6 min	Three pairs of copiers each differ in one setting. Learners run copier A and copier B, compare their shelves and choose why B forgot more: its own books, the tidy-up filter or small books.	• The tidy-up filter makes books look neater. Why does it make Quill forget faster? • How did you know which setting was the cause?	Only one setting differs, so it is the cause. Dropping unusual sentences throws away the rare words first.
5. Evaluate 5 min	Learners judge three apps: a homework helper, a translation app and a song maker. For each, they decide if it will lose rare things over time and back it with a fact from their runs.	• Why should the translation app stay healthy? • Which fact did you use for the song maker, and why does it fit?	Ask where the next training data comes from. Checked human writing keeps rare things; AI-made data and filters lose them.

Steps follow Bloom's revised taxonomy, in the order the Unit presents them.

Deepen the thinking

Critical thinking	Quill 8's sentences still look like sentences. What would you measure to show it has lost something?
AI thinking	Bigger books slowed the loss but did not stop it. Why can't more AI-made data replace real human writing?
Ethics and fairness	Real stories are slow to collect and robot books are free. Who should pay for fresh human writing, and should writers be asked first?
Connect to life	Have you seen a rumor or a meme change as people copied it? What got lost along the way?

Misconceptions to listen for

"More AI-made data is always more data."
Without fresh human writing the shelf only shrinks. Bigger books slow the loss but still leave rare words missing.

"Filtering out weird stuff makes the data better."
In Analyze, the copier with the tidy-up filter forgets faster. Unusual sentences are exactly where the rare words live.

"Lost words come back eventually."
Once no book contains a word, no later Quill can learn it. Only putting the students' stories back returns it.

In real life the card learners see at the end of the Unit

- More and more web pages, reviews and pictures are made by AI, and future AIs may learn from them.
- Translation and writing tools improve when real people check and correct them.
- A school library that only kept copies of copies would slowly lose its unusual books.

Try it this week. With an adult, find a web page you think an AI wrote. What would a model learn from it, and what would be missing?

Watch out for. AI-written text can look fine but leave out rare words, ideas and voices. Once a model stops seeing them, it stops writing them.

Talk about it. Whose stories might go missing if AIs mostly learn from other AIs?

Exit ticket

1. Which words does a copying robot forget first, and why?

2. Name one setting that slowed the loss and one that sped it up.

3. What keeps an AI from forgetting rare words over many generations?

Going further

To close the Module, the Data Explorer Lab lets learners clean a messy spreadsheet and hunt for patterns in sample data. For homework, learners pick an app they use and ask: where will its next training data come from?