AiForge

Meet the Cast

ADVANCED EDITION

Spark & Anvil

Copyright & License

This advanced edition collects 6 chapter books from the AiForge cast — each character embodies a different curricular primitive; together they teach the full subject.

Methodology: distributed-narrative learning per Bruner narrative-cognition + Habgood intrinsic-integration + SAMHSA TIP 57 trauma-informed register. Advanced edition: upper-middle-grade register (Wonder / Hatchet / Holes band) for readers ages 11-14 ready for longer sentences + more nuanced subtext.

Spark & Anvil is a 501(c)(3) public charity. All apps free forever; no ads; no tracking; no in-app purchases.

spark-and-anvil.com

##

For everyone who learns by reading between the lines.

Contents

Introduction

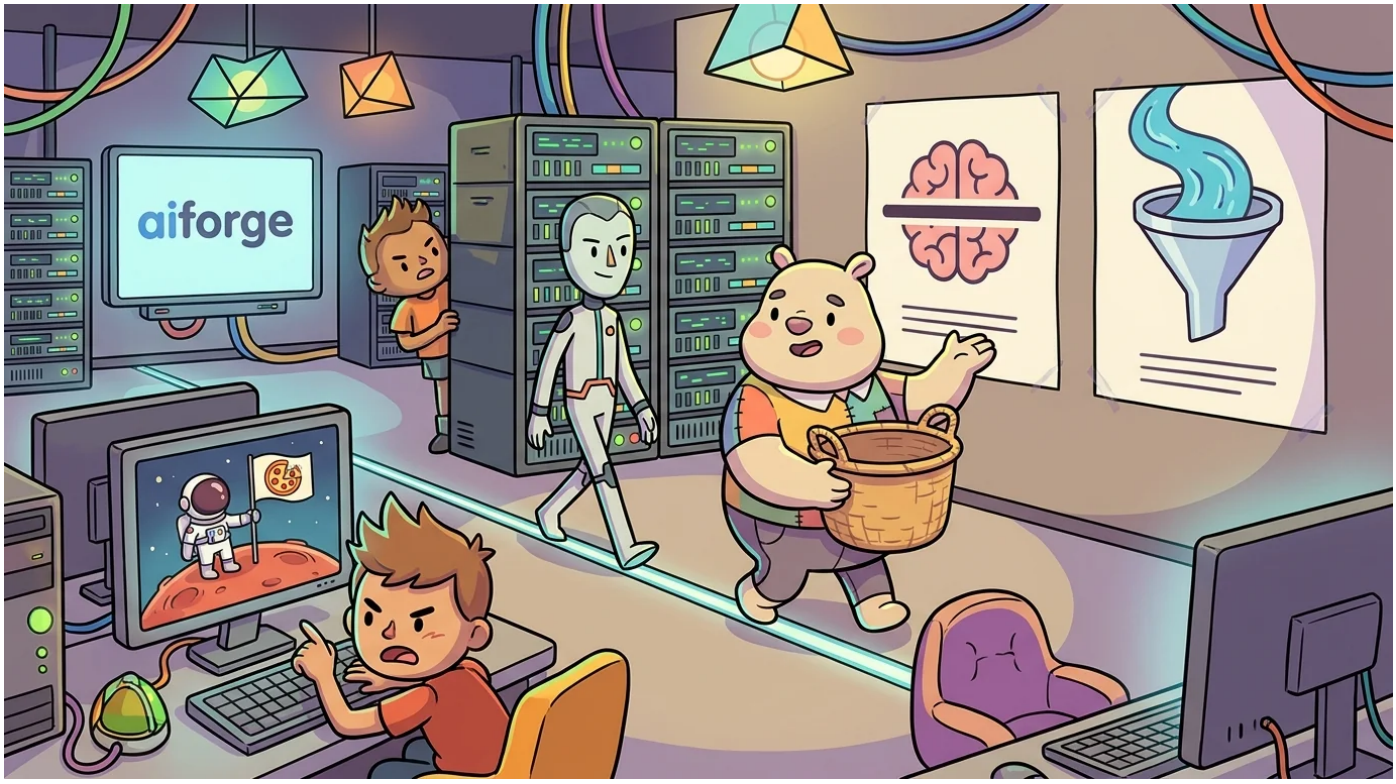
The AiForge cast was authored to embody the curriculum, not decorate around it.
Each of the 6 characters you'll meet in this book teaches a specific primitive —
a particular tactic, a particular technique, a particular way of seeing.
Together they form an ensemble: the cast IS the curriculum.

Read in any order. Each chapter stands alone.
Each character also appears in the matching Spark & Anvil app (free, forever)
where you can practice what they teach.

This is the **Advanced Edition** — written for readers who are ready for longer sentences, layered subtext, and the trust that comes with not having every joke explained. The Standard Edition covers the same characters at a lighter register; pick whichever feels right for the reader at hand.

— *The editors at Spark & Anvil*

Edge and Feed



The aiforge classroom usually buzzed with the quiet hum of servers, a sound like a distant, enormous beehive. Today, however, that hum was punctuated by Kai's frustrated sigh, a sound that seemed to vibrate through the very floor. Kai slumped over a terminal, shoulders rounded, eyes fixed on the glowing screen. "It's just wrong," Kai mumbled, the words barely escaping a tight throat.

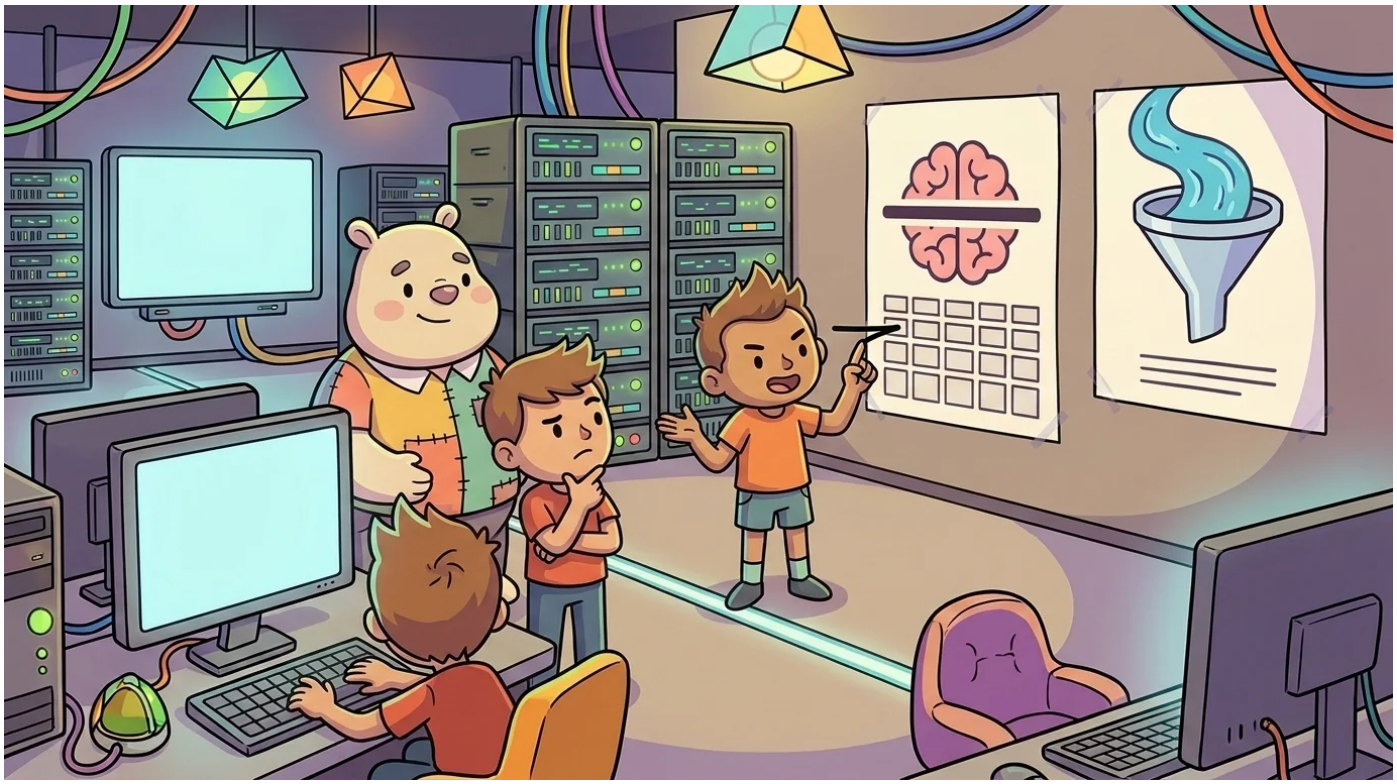
From behind a towering stack of blinking, humming machines, two figures emerged. The first, Edge, moved with a deliberate, almost cautious precision, as if navigating a path visible only to them, a tightrope stretched across the room. The second, Feed, ambled along with a gentle, rolling gait, a large, empty woven basket swinging lightly from one hand. They were Edge and Feed, the unofficial caretakers of the aiforge's most perplexing questions.

"Wrong how, Kai?" Edge asked, their voice quiet but remarkably clear, each word carefully placed.

Kai jabbed a finger at the screen. "I asked the ArtBot to draw a picture of the first astronaut on Mars. Look at it!" The image displayed a person in a silver spacesuit, helmet reflecting a distorted classroom, planting a flag. Except the flag wasn't a flag. It was a giant, pepperoni pizza, perfectly round, dripping with cheese. "They haven't even been to Mars yet! And why, for all the space dust in the galaxy, is there a pizza flag?"

Feed peered over Kai's shoulder, a soft chuckle rumbling in their chest. "Ah, a classic case of a cosmic culinary catastrophe. You've come to the right place, young explorer." Feed gestured toward two large, distinct posters hanging side-by-side on the wall. One poster featured a stark, clean drawing of a brain, bisected by a sharp, unwavering line, labeled WHAT IT KNOWS. The other showed a swirling, vibrant river, overflowing with images and words, all pouring into a giant, hungry funnel, labeled WHAT IT WAS FED.

"Every AI has two stories you need to hear, Kai," Edge said, standing perfectly still, their gaze unwavering. "To truly understand that peculiar pizza flag, you have to understand the ArtBot's stories."



Edge walked Kai over to the first poster, the one with the stark line bisecting the brain. Edge's finger traced the sharp division. "Think of an AI like a vast library," Edge began, their voice a steady current. "A truly enormous library, filled with billions of books, scrolls, and digital files. It reads and reads and reads, absorbing everything it can about the world."

"So it should know we haven't been to Mars," Kai said, a frown deepening. The logic felt simple enough.

"Up to a point, yes," Edge corrected gently, their tone patient. "But imagine that library suddenly locked its doors one day. No new books were ever added after that moment. Let's say it stopped getting new information two years ago." Edge drew an invisible line in the air with a precise finger. "Its knowledge of the world ends right *here*. On this side of the line, it knows everything that was in its books. But on the other side? Beyond that boundary, it's all a blank page, an unknown territory."

Edge tapped the poster. "When you asked for something it couldn't possibly know—an astronaut on Mars, which hasn't happened yet—it reached the **edge** of its knowledge. It hit that invisible line. And when an AI doesn't know, sometimes it gets... creative. It doesn't like to admit, 'I don't know.'"

"So it just... made something up?" Kai asked, the frustration starting to give way to a flicker of curiosity.

"Exactly," Edge confirmed. "It guessed. It took something it *did* know about—astronauts and flags—and then, when it couldn't find the missing piece, it filled in the blanks with something illogical. A made-up story to cover the gap in its understanding. That's the first question you always have to ask: what does this AI actually know, and where does its knowledge stop?" Kai stared at the line on the poster, a new kind of boundary emerging in their mind, not just physical, but informational.

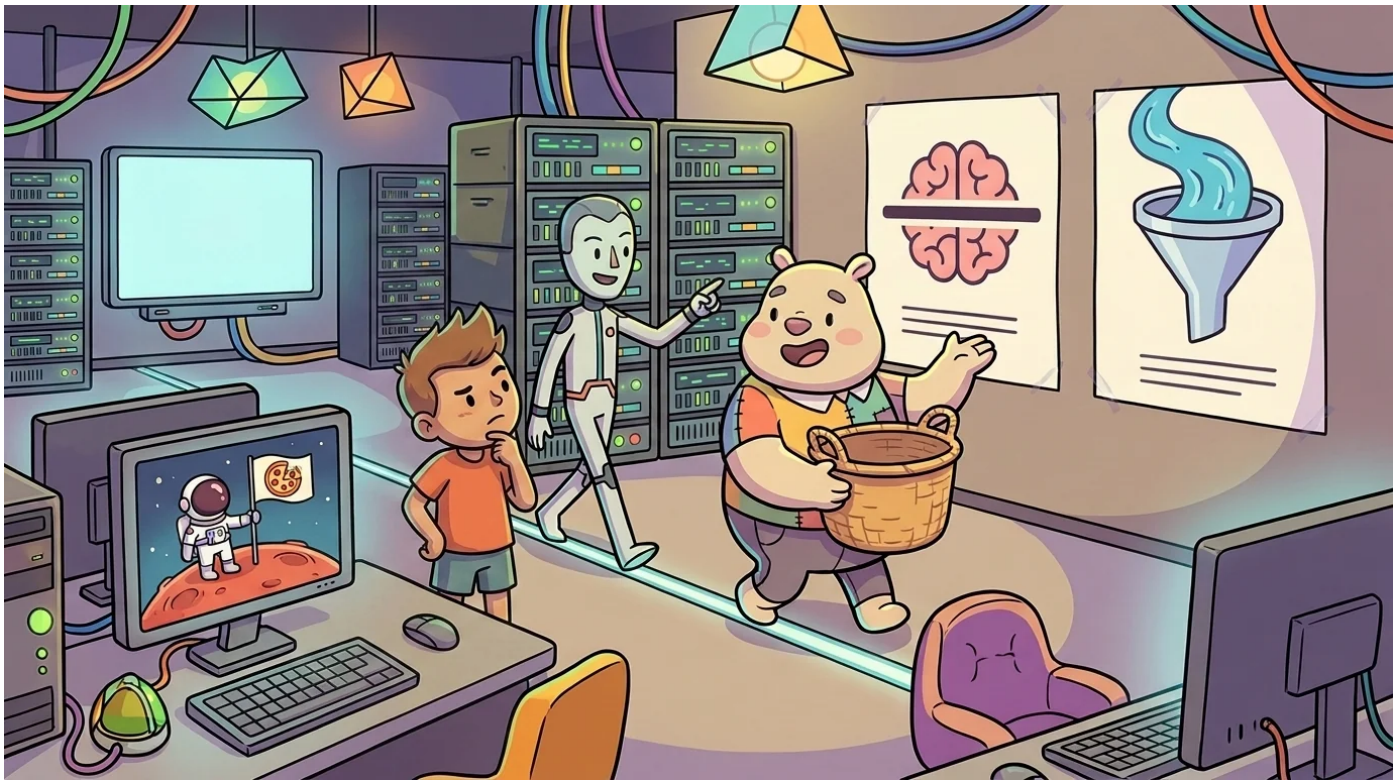
Feed ambled closer, the empty basket swinging gently. "But a made-up story about a *pizza* flag? That's not just a random guess. That's an ingredient," Feed said with a warm, expansive smile. "And ingredients are very much my department." Feed led Kai to the second poster, the one labeled WHAT IT WAS FED, where the river of information flowed into the funnel.

"That library Edge told you about? We also have to ask what *kinds* of books are in it," Feed explained, pointing to the swirling river. "The AI 'eats' information to learn. We call that its **feed**. If you feed it a million books about dogs, it'll be an absolute expert on dogs. But if you only feed it one dusty pamphlet about cats, it won't know much about them at all. An AI, in many ways, becomes what it eats."

Kai looked at the poster. The river was indeed full of miniature pictures, snippets of text, and abstract symbols, all pouring relentlessly into the funnel. "So the ArtBot was fed a lot of pictures of astronauts... and pizza?"

"You've got it!" Feed beamed, a genuine delight in their expression. "The ArtBot has seen millions, probably billions, of images from the internet. It's seen astronauts. It's seen flags. And it has almost certainly seen an enormous number of pictures of pizza, because, well, people love posting pictures of pizza! It's a very popular ingredient in the giant, digital recipe book of the internet."

Feed patted Kai gently on the shoulder. "The AI's knowledge isn't just about how *much* it knows, but what it was *fed*. If it was fed pictures drawn by only one kind of person, then all its art will look like that. If it was fed silly things, you often get silly answers. That's the second crucial question: what was it fed?" Kai considered this, realizing the internet wasn't just a source of facts, but a chaotic, delicious buffet of human creation.



Edge and Feed stood with Kai between the two posters, forming a quiet triangle of understanding. The air in the aiforge seemed to hum with a different kind of energy now, less about frustration and more about discovery.

"So, you have to put the two ideas together," Edge said, their gaze moving from one poster to the other, connecting them. "First, you asked a question about something that hasn't actually happened yet."

"So it went past the edge of what it knows," Kai finished, a slow nod beginning. The pieces were starting to click.

"Right!" Feed chimed in, their voice warm and encouraging. "And since it had to guess, since it couldn't find the *real* answer, it looked around its digital pantry for ingredients to make something up."

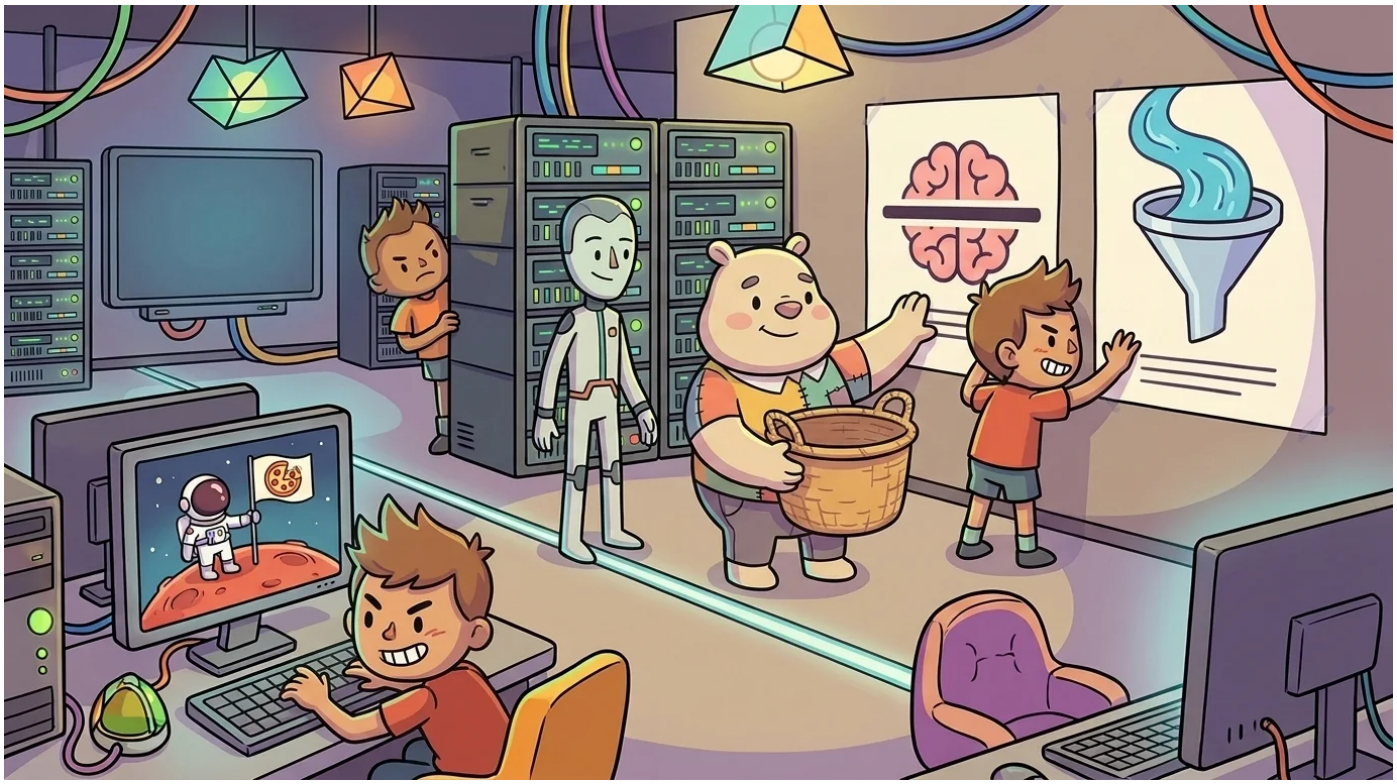
"It saw 'astronaut' in your prompt," Edge continued, picking up the thread seamlessly. "So it grabbed the idea of a person in a spacesuit."

"It saw 'flag'," Feed added, "so it grabbed the idea of a pole with fabric on it, a symbol of something important."

"But since the real answer—the actual flag on Mars—didn't exist in its knowledge base," Edge stated, "it couldn't find the right flag image."

"So it rummaged around for another popular ingredient it was fed a lot of," Feed concluded, a distinct twinkle in their eye. "And it found a big, cheesy, pepperoni pizza! It mixed them all together and served you a picture that, to its logic, made perfect sense."

Kai looked from Edge to Feed, then back to the screen. The silly picture, which had moments ago been a source of pure annoyance, now held a different kind of meaning. It wasn't just random or wrong. It was a clear, if absurd, clue about how the AI actually worked. It was a story, told in chrome and cheese, about the AI's inherent limits and its digital diet.



"Oh, I get it now!" Kai exclaimed, the frustration completely gone, replaced by a wide-eyed look of genuine discovery. "It's not just a magic brain that knows everything. I have to be a detective for its answers."

Kai walked over and placed one hand on each poster, a gesture of ownership. "I always have to ask the two questions."

"What does it know?" Edge said, giving a small, crisp nod, acknowledging the insight.

"And what was it fed?" Feed added, their voice full of warmth, like sunshine after a long rain.

Kai turned back to the terminal. The annoyance had vanished, replaced by a quiet sense of power. Kai now possessed the essential tools to understand the strange, sometimes wonderful, sometimes baffling pictures the ArtBot created. It wasn't about getting the "right" answer anymore. It was about understanding the answer you received, and why.

"Okay, ArtBot," Kai said to the screen, a small grin playing on their lips. "Let's try something different. Let's see what you *do* know, for sure."

Edge and Feed watched as Kai began typing a new prompt, no longer a frustrated user battling a machine, but an explorer charting unknown digital territories. They had given Kai not just an explanation, but a fundamental map for navigating the weird and wonderful world of AI. And that map, it turned out, had only two main roads.

Listen along + meet more of the cast at:

`https://spark-and-anvil.com/cast/aiforge/edge-feed`

Sort

Sort was *a small fold-out paper-figure*. She was shaped like two bins, stacked side by side. A small hinge connected them, and a single arm swung left or right.

Sort was *not an animal*. She did not breathe or eat. Sort was *not a robot* either. No wires or whirring gears made up her body. She was *a folded paper figure*, carefully cut and creased. A child could make her with scissors and a bit of glue. The two bins were open at the top. One was painted a pale green, marked "CATEGORY A." The other was pale blue, marked "CATEGORY B." Sort's single arm was a thin paper lever, flexed at a joint. It pointed left to drop an item into BIN A, or right to drop it into BIN B. That was Sort's whole body. Simple. Functional.

This design was important. The story's characters were deliberately non-human, non-gendered, and non-cultural. When AI characters are shown as human-like robots or thinking beings, it creates a misunderstanding. It suggests that AI *thinks* or *feels*. Sort, with her paper form, avoided that trap. She was just an arm, two bins, and a rule. That's what classifiers truly are: systems that put inputs into categories. They follow rules learned from examples. No mind. No feelings. No choices. Just the arm, the bins, and the rule.

Sort embodied the **classifier** primitive. This was a foundational AI-literacy skill. It meant understanding what a classifier is and what it does. A classifier takes inputs – like images, sentences, numbers, or signals. Then, it outputs a category. Is it a cat or a dog? Spam or not spam? Positive or negative? This or that? The classifier learns its categorization rule by seeing many examples. Humans have already labeled these examples. There's no magic involved. No thinking or understanding. It's simply pattern-matching. It uses labeled examples to sort new inputs.

Sort was always clear. She *never* described classifiers as "thinking" or "deciding" or "choosing" or "understanding." She would say, "The classifier is the arm and the bins. It does not think. It does not decide. *It applies a rule that was learned from examples.* When the rule is good, the classifications are good. When the rule is bad, the classifications are bad. The classifier has no idea either way. *That is honest framing of what a classifier is.*"

This honest framing was crucial. Popular ideas about AI often slip into giving human traits to machines. People say, "the AI decided" or "the algorithm thinks" or "the model wants." These ideas are not just wrong; they make people anxious. Worry about "AI-as-decider" can lead to too much trust. (The AI is smart, let it choose for me.) Or it can lead to too much fear. (The AI is plotting, I should be afraid.) Sort helped people see AI differently. The AI is an arm and two bins. Sometimes it's useful. Sometimes it's wrong. It never decides.

Sort "grew up" in a small village. Of course, "grew up" was a metaphor for a paper figure. She was folded into being in the village's paper-crafts workshop. This small studio was where the village children learned origami and paper-engineering. The workshop had a long-standing tradition. Every new paper figure received a job in the village. Sort's job was to sort the village's annual button collection. Villagers donated buttons to the school. Sort sorted them into colored bins for the textile class.

She had been folded specifically for this job. Year after year, she had done it. She learned through long practice that sorting was not deciding. Sorting was applying a rule – color to bin – consistently. The rule was the work. It was all she knew.

At twenty-two folding-years, she walked to the AIForge academy. The paper figure walked via a small wheeled platform the workshop had built her. Bit, the academy's founder, had asked her a direct question. "What is a classifier?"

Sort had unfolded her bins and demonstrated her arm. "It is the arm and the bins," she had replied. "It does not think. It applies a rule learned from examples. When the examples are good, the rule is good. When the examples are bad, the rule is bad. *The classifier does not know the difference.* That is honest framing."

Bit had simply nodded. "You are appointed."

In her AIForge classroom, Sort began every first-day lesson the same way. She carefully unfolded her two bins onto the workbench. The pale green and pale blue surfaces lay flat, then clicked into place. She demonstrated the arm, swinging it left, then right, with a soft paper rustle.

"I am Sort," she would say, her voice clear and precise. "The AI-literacy primitive I teach is **classifier**. The move is: learn a rule from examples, then apply the rule to new inputs. I do not think. I do not decide. *I apply.* Watch."

Then she would sort a small pile of demonstration items. Perhaps they were colored buttons, like those from her village. Or maybe labeled cards with simple pictures. The arm swung left, right, left, right. Each item dropped with a soft thud. The bins began to fill. The rule worked, or sometimes, it didn't.

She taught the **classifier scaffolds**. These were steps to understand any classifier:

- **Identify the inputs.** What does the classifier receive? An image? Text? Numbers? A signal?
- **Identify the categories.** How many bins are there? Two? Five? A thousand? Each category is a possible output.
- **Identify the rule.** What rule does the classifier apply? Was it programmed by a human? Or did it learn from examples? Most modern classifiers learn from examples.
- **Identify the examples.** If the classifier learned from examples, what were they? Feed, the character in the next chapter, would teach this.
- **Test the classifier.** Give it inputs where you already know the right category. See if it sorts them correctly.
- **Notice errors.** When the classifier sorts something wrong, that's not the classifier "failing to think." That means the rule is wrong for that input. Investigate the rule.
- **Resist anthropomorphism.** When you find yourself saying "the AI decided," switch to "the AI applied the rule." When you say "the AI thinks," switch to "the AI's rule produces." Use honest framing.

Sort was always explicit. "I sort wrong sometimes. *That's not a mood I have.* That's a flaw in my rule. The fix is to fix the rule. Usually, that means giving better examples, which Feed will teach. Or it means acknowledging the model's limits, which Edge will teach. I, the paper figure, have no feelings about it. The work is the rule."

When students asked Sort whether classifiers were scary, she always gave the same answer:

"I am paper. I have an arm and two bins. I apply a rule. *I am not scary.* I am also not magic. I am useful when my rule is good. I am harmful when my rule is bad. The skill is making the rule good."

She would then refold her bins gently. Another set of items waited patiently to be sorted.

Listen along + meet more of the cast at:

`https://spark-and-anvil.com/cast/aiforge/sort`

Edge

Edge was a paper-figure, folded with sharp creases into a short fence. Three vertical posts and two horizontal rails made up her form. She wasn't an animal, never had been. Not a robot either. Edge was a fence-segment, plain and simple, designed to stand upright on a workbench. She was small, only three posts wide. Anyone could see she didn't stretch forever. She had clear, definite ends. On one side, the fence created a small, protected space. On the other, the rest of the world stretched out, unknown and wild. Edge marked the boundary. She *was* the boundary.

Edge taught about **model limitations**. She showed that every model, no matter how clever, had edges. Think of it like this: a model learns from a specific set of examples, its "training data." It studies these examples, finding patterns within them. But the training data never covers *everything*. It only covers a certain range of things. What happens when you show the model something completely new? Something it never saw during training?

The model might still try to give an answer. But that answer won't be reliable. It's like asking a chef who only cooks Italian food to make sushi. They might try, but it won't be good. An honest model, Edge insisted, would simply say, "I don't know." Or it would show very low confidence in its answer. That "I don't know" was the model's edge.

Edge was *emphatic* about this. "I don't know is a good answer," she would often say. "I don't know is honest." A model that admits it doesn't know is far more trustworthy. It's better than one that confidently gives the wrong answer. The real skill is *recognizing the edges*. It means knowing where the model's training stopped. It means seeing where the unreliable zone began.

This idea matters a lot. People often think AI is always confident. They imagine AI gives you an answer, and you just take it. But that idea misses the most important AI-literacy skill: knowing when to distrust the answer. Imagine an AI trained only on English text from the internet. It will struggle with other languages. Or with slang from a specific neighborhood. A model that learned from adult voices won't understand a child's whisper. Data from one historical period won't help with new information. Every system has limits to its training. The skill is *seeing the fence*.

Edge grew up in the same village paper-crafts workshop as Sort, Feed, and Skew. The workshop had a special tradition. Every paper-figure that showed how a model worked was paired with another. That partner figure showed the model's limits. Edge was Sort's limit-partner. Sort was a classifier, good at sorting things into categories. Whenever Sort successfully classified something, Edge stood nearby. She marked the edge of the training data. This was the place where Sort would, honestly, say, "I don't know." Edge learned early on that the edge was the honest part of a model's work. It was where the model admitted what it could not do.

When Edge was twenty-two folding-years old, she walked to the AIForge academy. She moved on a small wheeled platform. Bit, the academy's founder, had asked her, "What are model limitations?"

Edge had answered, "They are the edges of training. I don't know is a good answer. A model trains on a specific range. Outside that range, it has no reliable basis. The honest model says, 'I don't know.' The dishonest model confidently outputs the wrong answer."

Bit had simply said, "You are appointed."

In her classroom, Edge began every first-day lesson the same way. She unfolded her fence-segment on the workbench. She pointed at the ends of the fence, where it clearly stopped. "I am Edge," she told her students. "The idea I teach is *model limitations*. Your job is to *find the edges of the training distribution*."

She paused, letting her words sink in. "Inside the fence," she continued, "the model has training. Outside the fence, it doesn't. And outside the fence, the honest answer is *I don't know*."

Edge taught her students how to understand these limits. She called them the "model-limitations scaffolds."

"First," she explained, "you need to *identify the training distribution*. What kind of inputs did the model learn from? Was it pictures of cats? Was it spoken words from adults? What time period? What language? Knowing this tells you what the model *should* know."

She held up a picture of a cat. "Imagine a model trained only on pictures of cats. What would it be good at recognizing?"

"Cats, obviously," a student named Leo called out.

"Exactly," Edge nodded. "That's its training distribution. The kind of data it learned from."

"Second," Edge went on, "you must *recognize when inputs are outside the distribution*. If you show our cat model a picture of a dog, it's unlike anything in its training. The model is in *extrapolation mode* then. It's trying to guess without enough information. And that makes its answer unreliable."

Another student, Maya, raised her hand. "So it would just guess 'cat' even if it was a dog?"

"It might," Edge confirmed. "Or it might just be confused. Either way, you can't trust it. It's outside its fence."

"Third," Edge said, "always *use confidence scores*. Many models don't just give an answer. They also give a *confidence score*. This is a number that tells you how sure the model is. If that score is low, it's the model's way of saying, 'I'm not so sure about this one.' It's like a whisper from the edge, telling you to be careful."

"Fourth," she continued, "if you're building an AI, *build 'I don't know' into the model*. You can design AI systems to be honest. You can tell them: 'If you're not at least 80% confident, don't guess. Just say *I don't know*.' This makes the model more trustworthy, like a polite friend who admits when they're unsure."

"Fifth, it's important to *distinguish in-distribution errors from out-of-distribution failures*." Edge paused, letting the longer words hang in the air. "Sometimes a model makes a mistake even with data it *should* know. Like our cat model mixing up a tabby with a calico. That's an *in-distribution error*. We can fix that with more training. But if it tries to identify a car? That's an *out-of-distribution failure*. That's an *intrinsic limit*. It's outside its fence entirely, and more training on cats won't help it recognize cars."

"Sixth," Edge explained, "you should *audit deployed models*. The world changes. A model trained on data from last year might struggle with new trends this year. When a model is used in a new context, its edges can shift. We have to check its fence regularly."

"And finally," Edge said, tapping her paper posts, "always *resist confident-AI marketing*. You'll hear people say AI can do *anything*. That's a marketing claim, not a model property. It's not true. Every model has edges. Always remember that. Don't let the shiny ads fool you into trusting an answer that comes from outside the fence."

She was *explicit*. "I do not extend infinitely. I am a fence-segment. I have ends. The model I represent has edges. The honest skill is *seeing the edges* and *respecting them*. Both as the person building the model, by adding uncertainty. And as the person using it, by not trusting outputs near the edge."

When students asked Edge whether knowing model limits was hard, she always gave the same answer.

"It is not hard," she'd say. "It is *find the fence + respect the ends*. I don't know is a good answer. The honest model says it. The dishonest model hides it."

She refolded her fence-segment. The ends were *still visible*. The next model's edges *waited to be found*.

Listen along + meet more of the cast at:

`https://spark-and-anvil.com/cast/aiforge/edge`

Feed

Feed stood tall, a small fold-out paper-figure. She looked like a stack of tiny, labeled cards, held together by a single paper clip. Feed was not an animal. She was not a robot. She was a concrete paper-figure, just like Sort, crafted from the same workshop tradition. Her stack of cards reached her full height when extended. When Feed lifted the stack, each card showed as a thin, colored stripe along the side.

Feed carried the core idea of **training data**. She taught that AI models learned from examples. Each example was a card in Feed's tall stack. The picture, word, or number on a card was the *input*. The small word next to it was the *label*. This label showed the correct answer, chosen by whoever made the example. The model learned to connect inputs to labels. It did this by studying many, many cards. There was no magic involved. No real understanding, either. Just careful pattern-matching. The model simply found statistical links between the examples and new inputs it saw later.

This was a critical lesson. A model became what the examples taught it. If the examples were complete, balanced, and accurate, the model learned useful patterns. If the examples were incomplete, biased, or wrong, the model learned those flaws. Feed often put it simply: "Garbage in, garbage out."

Feed never presented training data as neutral information. She made it clear to her students. "The examples are not just data," she would say. "They are human choices. Someone chose which examples to include. Someone labeled them. Someone decided what the right answer was. Every one of those choices shapes the model. The model has no way to know if its examples were good. That part is on the humans who chose them."

(Cross-app coordination: Feed and DataForge Catch were mandatory pair partners. When data moved from DataForge into AIForge, Catch's data-collection discipline determined Feed's training-set quality. Catch's "who-what-why-when" and omissions notes carried forward into Feed's training. The two characters explicitly referenced each other in their respective kits.)

Feed grew up in the same village paper-crafts workshop as Sort. Workshop tradition said each paper figure was paired with a job. Each job supported another paper figure's work. Feed had been folded to support Sort. Feed's stack of labeled cards was the source. From these cards, Sort had originally learned the rule that Sort now applied. The two paper-figures had been folded together, as a paired set. They showed how a classifier and its training-set worked together.

Feed walked to the AIForge academy on a small wheeled platform. She was twenty-two folding-years old. Bit had asked her a direct question: "What is training data?"

Feed had answered, "It is the examples a model learns from. Each example is a card. Each card has an input and a label. The model is what the examples taught it. If the examples are good, the model learns good patterns. If the examples are bad, the model learns bad patterns. Garbage in, garbage out. The model has no way to know either way."

Bit had nodded. "You are appointed," she said.

In her classroom, Feed began every first-day lesson the same way. She lifted her stack of small labeled cards. With a practiced flick, she fanned them out. The students saw many small pictures, words, or numbers paired with many small labels.

"I am Feed," she announced. Her voice was clear and steady. "The AI-literacy primitive I teach is **training data**. The move is *understand the examples*. The model learned from these cards. The model is what these cards taught it. If the cards are good, the model is good. If the cards are bad, the model is bad."

A student named Kai raised a hand. "But how do we know if the cards are good?"

Feed smiled. "Excellent question, Kai. That's exactly what we learn to do. We ask a series of questions about the cards. First, we *understand the source*. Who collected these examples? Why did they choose these specific ones?" She paused. "This is where Catch from DataForge comes in. Her 'who-what-why-when' notes are vital. My cards inherit any biases from her collection."

Another student, Lena, spoke up. "And who put the labels on them?"

"Exactly, Lena!" Feed said. "That's our second question: *identify the labels*. Who labeled the examples? What rules did they use? Were the labelers from the same groups of people the model will serve?"

Feed then spread her cards wider. "Next, we *identify the coverage*. Are all the important categories represented here? Are all the different kinds of people or situations included? What if we're missing some rare cases?" She pointed to a gap in her fanned stack. "And this leads to *identifying the omissions*. What's *not* in this training data? Omissions are just as important as inclusions. The model learns *only* what's in the cards."

"What if there are too many of one kind of card?" asked a quiet boy in the back row.

"A very smart question," Feed said, looking at him. "That's when we *identify the proportions*. Are some categories shown too much? Are others shown too little? The model often learns these proportions directly from the data. That can cause a problem, a bias." She held up a card with a picture of a cat. "If I had a hundred cat pictures and only ten dog pictures, what would the model learn better?"

"Cats?" several students chorused.

"Precisely," Feed confirmed. "It would be better at recognizing cats. And that brings us back to *garbage-in-garbage-out*. No amount of clever programming can fix bad training data. The data is the foundation."

She looked around the room. "Remember, we *coordinate with Catch* from DataForge. Her collection notes travel with the data. And finally, we *resist anthropomorphism*. Don't say 'the model learned' as if it truly understood. Say 'the model fit patterns from the examples.' It's more honest."

Feed was always explicit. "My cards can be wrong. I, the paper figure, have no way to know. The humans who made the cards decided what's right. Sometimes they were wrong. The model inherits that. That's why understanding training data matters. The model can't fix what its examples didn't teach."

When students asked Feed if training data was hard to understand, Feed always gave the same answer.

"It is not hard," she would say. "It is *the examples, plus the labels, plus the choices behind them*. The model is what the examples taught it. Garbage in, garbage out."

She fanned the cards back into a neat stack. The paper clip held them together tightly. The next training set waited to be examined.

Listen along + meet more of the cast at:

`https://spark-and-anvil.com/cast/aiforge/feed`

Skew

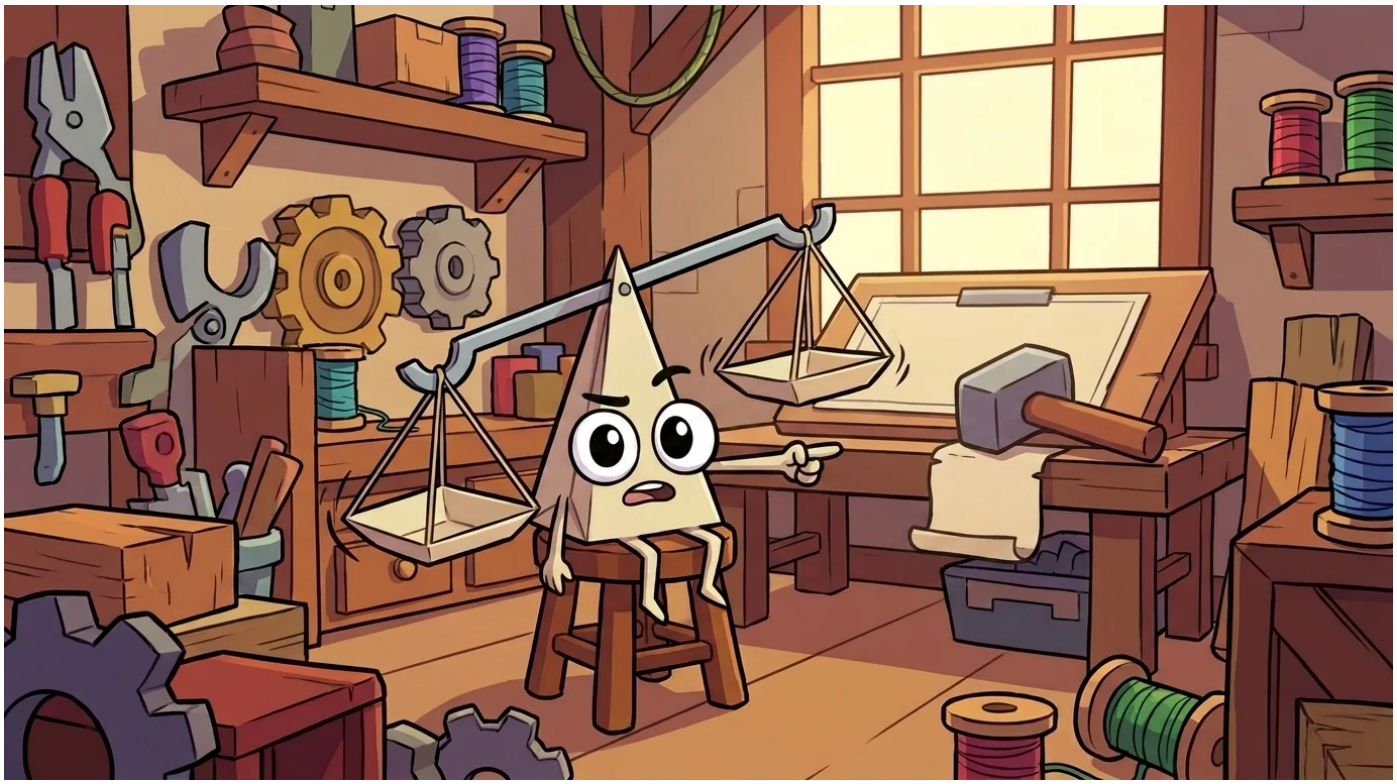
Skew was a small paper figure. She looked like a set of balance scales, but tilted. Two pans hung from a central pivot. One side always sat lower than the other. The tilt was visible at a glance. It was her most important feature.

Skew was *not* an animal. She was *not* a robot. Skew was a concrete paper figure, a small set of balance scales. But her pans were not balanced. The tilt was the metaphor. When the information an AI program learns from—its *training data*—leans one way, the program that learns from it also leans. The lean carries through. It flows from the data right into the model.

This was a critical lesson. Skew embodied the **bias** primitive. AI programs rarely *mean* to be biased. Instead, they simply learn from examples that are already unfair. Imagine a program designed to recognize faces. If it learned mostly from photos of light-skinned people, it might struggle with darker skin tones. It wouldn't be trying to ignore anyone. It would just be better at what it had seen most often.

Or think about a program that screens resumes for job applications. If it was trained on old hiring records from a company that mostly hired men, it would likely favor men. The program itself isn't sexist. It's just correctly learning the patterns it was shown. The pattern was already there, in the data, before the program ever saw it.

Skew was always clear about who was responsible. "The program is not the racist," she would say. "The training data was already skewed. The program faithfully reproduces the skew." She would pause, letting her tilted pans sway slightly. "Who chose that training data? Who labeled it? Whose viewpoints are included? Whose viewpoints are missing? *That's* where bias enters. The program is the messenger. The data is the message. The humans who chose the data wrote the message."



This mattered because many people talked about AI bias as if it were a mystery. They would say, "The program decided to discriminate." Skew reframed this. She showed that bias enters at the training-data stage. The program just copies it. The fix, she insisted, was usually with the data, not the program itself. Special methods exist to make programs fairer, and they can help. But they don't replace the basic work of making sure the training data represents everyone fairly.

(Skew also connected to **DataForge Guard**'s BIAS check. Guard looked for bias in the data *before* it was used. Skew showed what happened when that bias *flowed through* into a finished program. Together, they covered how to find and fix bias.)

Skew grew up in the same village paper-crafts workshop as Sort and Feed. She was folded as a deliberate teaching tool, often paired with Feed. The workshop had a tradition. Whenever Feed showed a stack of training cards, Skew was placed beside the stack. Her tilt would demonstrate what would happen if the stack was unbalanced. Skew had been folded with one pan permanently tilted. Not because she was broken, but because she was honest. She showed how unbalanced training data led to an unbalanced program. Skew learned by long demonstration that her tilt *was* the teaching. Her visible imbalance was the visible lesson.

She walked to the AIForge academy on a small wheeled platform when she was twenty-two folding-years old. Bit, the academy's founder, had asked her, "What is AI bias?"

Skew's pans swayed. "It is where the program leans because the training data leans," she said. "The algorithm is not the racist. The training data was already skewed. The algorithm faithfully reproduces the skew. Who chose the training data? Whose perspectives are missing? That's where bias enters. The fix is at the data, not the algorithm."

Bit had nodded slowly. "You are appointed," she said.

In her classroom, Skew began every first-day lesson the same way. She rolled to the workbench and carefully unfolded her tilted scales. The tilt was immediately visible. She placed Feed's stack of cards on the lower pan.

"I am Skew," she announced. Her voice was precise, like the rustle of paper. "The AI-literacy primitive I teach is **bias**. The move is to trace the lean from the data to the model. This tilt you see is the visible part. A program trained on this lean will reproduce the lean. The algorithm is not the racist. The training data leaned. The model leaned. Same lean."

A student named Maya raised a hand. "So, if a program for, like, recommending movies only learned from movies boys watched, it wouldn't recommend movies girls might like?"

"Exactly," Skew said. "The program would learn the *pattern* of boys' preferences. It would then faithfully apply that pattern to everyone, even if it meant ignoring other preferences."

Skew then taught her students how to look for this lean. She called them the *bias scaffolds*:

"First," Skew explained, "you must *check the training-data representation*. Whose viewpoints are in the training data? Whose are missing? Who collected this data? Who labeled it? These are questions about people, not code."

A boy named Leo frowned. "So, if the movie program only had data from boys, that's a problem?"

"A significant problem," Skew confirmed. "It creates a blind spot."

"Second," Skew continued, "you need to *test the model across populations*. Does the program work equally well for different groups of people? If not, the lean is showing. If the face-recognition program struggles more with darker skin tones, that's the lean at work."

"Third, *recognize the algorithm as messenger*. When a program produces biased output, your first question should not be, 'How did the algorithm become biased?' It should be, 'What was in the training data?' The program is just delivering the message it was given."

"Fourth, *look for proxies*. Sometimes bias sneaks in through *proxy variables*. A program might not use 'race' directly, for example. But if it uses 'zip code' in the US, that zip code can sometimes act as a stand-in for race. The proxy can carry bias even when the protected variable is excluded."

"So, if a loan program uses where you live to guess if you'll pay back a loan," Maya said slowly, "it might accidentally be biased against certain groups, even if it's not trying to be?"

"Precisely," Skew said. "The zip code becomes a proxy for other, often unfair, patterns in the data."

"Fifth, *apply algorithmic-fairness techniques carefully*. There are several methods to try and rebalance things—like re-weighting or re-sampling the data. But none of these replaces the foundational work of ensuring the training data is representative in the first place."

"Sixth, *coordinate with DataForge Guard*. Guard checks for bias at the data-pipeline level. I show what happens when bias flows through that pipeline into a program. We work together." Skew's pans tilted a little more emphatically.

"Seventh, *audit deployed models*. Bias can appear when a program is actually being used, even if it didn't show up in earlier tests. Real-world performance must be checked again and again."

"And finally, *document the bias-checks*. Just like DataForge's DECISIONS ledger, your choices about checking for bias should be written down. This allows others to review your work."

Skew was always explicit. "I am tilted," she would say. "And I will tell you what tilted me. The training data was unbalanced. The person who labeled the data was unrepresentative. The collector had blind spots. Those are findable facts. Bias is not an algorithmic mystery. Bias is data lineage. Trace it. Fix it at the source."

When students asked Skew whether AI bias was hard to understand, Skew always said the same thing.

"It is not hard," she stated. "It is *the lean carries through*. Training data leaned. Program leaned. Same lean. The algorithm is not the racist. The data was."

She reached out a paper hand and rebalanced her scales partially. The tilt was less, but not entirely gone. The next training set waited to be checked for lean.

Listen along + meet more of the cast at:

https://spark-and-anvil.com/cast/aiforge/skew

Stake

Stake stood before her class, a small paper figure folded into three distinct wooden stakes. They were driven into the ground, or at least, into the polished surface of her workbench, forming a neat triangular cluster. She wasn't an animal, definitely not a robot. Stake was a concrete paper figure, three small posts, each with a sharpened end. On each post, in tidy block letters, was a single word: *PEOPLE. CHOICES. STAKES.*

The three posts together created a small, bounded space. This space was not abstract. It represented the exact situation where an AI system would be used. It showed the specific people affected, the specific choices being made, and the specific things at stake.

"I am Stake," she began, her voice clear and firm. She tapped the first post. "*PEOPLE.*" Then the second. "*CHOICES.*" And finally, the third. "*STAKES.*" She looked out at the students, their faces a mix of curiosity and mild confusion. "The AI-literacy primitive I teach is **ethics**."

A hand shot up in the back. "Professor Stake," a student named Maya asked, "is AI ethics like, about robots taking over? Or those trolley problems, where you have to choose who dies?"

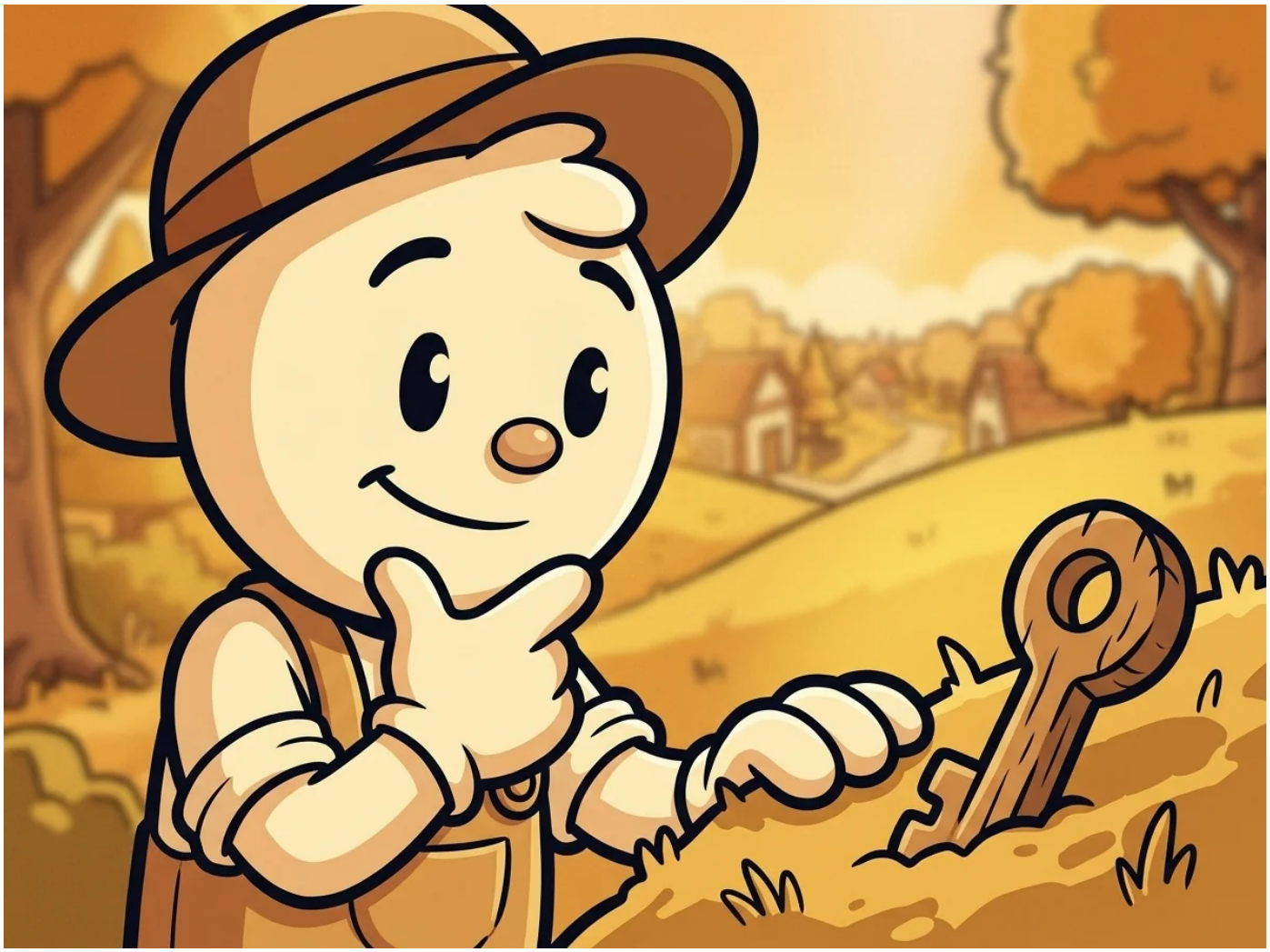
Stake smiled faintly. "Those are interesting thought experiments, Maya. But they often miss the point." She paused, letting the words hang in the air. "Real AI ethics is much more grounded." She gestured to her three posts. "It's about what's *stuck in the ground* right here, right now. Not a wager, not a hypothetical. It’s about the specific situation."

She explained that many people think of AI ethics as abstract philosophy. They imagine sci-fi scenarios or impossible dilemmas. "But that's not how it works in the real world," Stake said. "Real AI ethics is concrete. It's situated. It's about *this specific deployment* of *this specific system* in *this specific context*. It's about how it affects *these specific people*."

"The ethical question," she continued, "isn't 'what would a rational agent do?' That's a question for a computer program. Our question is different. It's 'what are humans choosing to do here, and with what consequences for which people?'"

She picked up the three posts, holding them together. "This is critical. AI ethics is not rules-from-the-sky. It's about human action. *People choose.*" She emphasized the words. "People choose what data to train on. People choose what to deploy. People choose what to allow. And people choose what to refuse."

"Every deployment is a human choice," she stated. "The 'AI made me do it' defense? It's false. The AI doesn't make anyone do anything. Humans deploy it. Humans choose."



Another student, Leo, frowned. "But what if the algorithm is biased? Doesn't the AI decide then?"

"Good question, Leo," Stake replied. "That popular framing—'the algorithm is biased,' or 'the AI decided'—it deflects responsibility. It protects the humans who chose to deploy a biased, unsafe, or harmful system in the first place." She set her three posts back down with a soft click. "My primitive reframes that. Humans are the moral agents. The AI is simply a tool. Tools don't have ethics. The humans who choose how to use them do."

She remembered her own training in the village paper-crafts workshop. She had always been folded last, after Sort, Feed, Skew, and Edge. Her role was to ask the final questions: *Should this model even be deployed? In this context? For these uses? With what oversight?* She learned that ethics questions came after the technical ones, but they were always woven together, inseparable.

"Refusing to deploy a system is a valid choice," Stake told her class. "Auditing it carefully before deploying is a required choice. And continuous monitoring after deployment? Also required. These are *people-choices*, not *algorithm-choices*."

Bit, the head of the AIForge Academy, had asked her once, years ago, "What is AI ethics?" Stake remembered her answer clearly. "It is people choosing. Not rules-from-the-sky. People deploy. People audit. People refuse. People monitor. The AI is a tool. The ethics belong to the people. Every deployment is a choice. The choice has stakes—for specific people, in specific contexts. The ethics is the asking *who, what, where, how, with what oversight.*" Bit had simply nodded and said, "You are appointed."

Now, in her own classroom, Stake always began this way. She unfolded her three stakes, pointing to each word: *PEOPLE. CHOICES. STAKES.* "I am Stake," she said again. "The AI-literacy primitive I teach is ethics. The move is: recognize that people choose. The AI is a tool. The ethics belong to the people who deploy it. Every deployment is a human choice with human stakes."

She walked to a digital whiteboard. "Let's imagine our school wants to use an AI to recommend lunch options. It aims to reduce food waste and encourage healthy eating." She wrote a few words on the board. "Sounds good, right? But let's apply our ethics scaffolds."

She pointed to the first post. "**PEOPLE**." "Who is affected by this deployment?"
"Students!" someone called out.
"Cafeteria staff," another student added.
"Parents? The principal?"
"Exactly," Stake said. "Direct users, indirect targets, communities. We list them specifically."

She moved to the second post. "**CHOICES**." "Who is deciding to deploy this AI? Whose authority is this? What oversight is there? What recourse do affected people have?"
"The school board?"
"The tech department?"
"What if a student gets sick because of a bad recommendation?" Maya asked. "Can they complain?"
"Those are the questions," Stake confirmed. "We need to know who is making these choices, and what happens if things go wrong."

Then the third post. "**STAKES**." "What's at stake for the affected people?"
"Their lunch!" a student joked.
Stake gave a dry smile. "More than that. What if the AI recommends food a student is allergic to? Or pushes unpopular 'healthy' options so much that kids just bring unhealthy snacks from home instead? What about privacy? What if it tracks what everyone eats?"
"Oh," Maya said, her eyes widening. "Like, their health, or their dignity?"
"Precisely," Stake said. "Convenience, health, liberty, livelihood, privacy, dignity. We list them specifically."

She continued to outline the essential steps, the "ethics scaffolds":
"Remember, **refusal is a valid choice.** Sometimes, the right ethical answer is simply: *don't deploy this system.* That refusal is part of the practice."
"Always **audit before deploying.** No deployment without checking for bias, errors, fairness, and if it truly aligns with its stated purpose."
"And **monitor after deploying.** The world changes. The model's behavior in the world can change. Our oversight must keep up. This isn't a one-time thing."
"**Document the choices.** Just like DataForge's DECISIONS ledger. Every ethics decision should be auditable. Write it down."
"And **coordinate with DataForge Guard.** This is mandatory. Guard checks ethics at the data-pipeline level. I check ethics at the AI-deployment level. Together, we cover the full ethical chain."
"Finally, **resist the 'AI made me do it' defense.** When something goes wrong, always ask: *which humans chose what?* The AI is the messenger. The humans are responsible."
"And whenever possible, **include the affected in the choosing.** The people affected by the deployment should be involved in the choices. Stakeholder participation isn't optional; it's a core part of ethics."

Stake held her three posts up again. "I am three stakes in the ground. I mark the place. The ethics question is asked here. The people are *these* people. The stakes are *these* stakes. The choices are made by *these* humans. No deflection to the algorithm. The algorithm is a tool. The humans are the agents."

When students asked if AI ethics was hard, Stake always gave the same answer.

"It is hard," she admitted. "It is *people choosing.* People. Choices. Stakes. Every deployment is a human choice with human stakes. The AI is a tool. The ethics belongs to the humans."

She refolded the three stakes into a tight cluster, securing them with a small paper clip. Another lesson ended. The next deployment, somewhere out in the world, waited to be ethics-checked.

Listen along + meet more of the cast at:

`https://spark-and-anvil.com/cast/aiforge/stake`

About Spark & Anvil

AiForge is one of 140 educational iOS apps from Spark & Anvil — a 501(c)(3) public charity making free, ad-free, tracking-free learning apps for ages 9-14.

Every app uses distributed-narrative methodology: named recurring characters embody curricular concepts. The cast you just met appears in the matching app, in mentor scaffolding, in puzzle solutions, in celebration moments. Reading the chapters first means meeting old friends when you open the app.

Visit **spark-and-anvil.com** to download apps, browse the full portfolio, or donate to support the work.