AiForge

Meet the Cast

STANDARD EDITION

Spark & Anvil

Copyright & License

This book collects 5 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.

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 hearing a story first.

Contents

- Copyright & License
- Contents
- Introduction

Edge

- Voice register
- Arc across kits
- Relationships
- Cultural-sensitivity gate
- Cultural-context note

Feed

- Voice register
- Arc across kits
- Relationships
- Cultural-sensitivity gate
- Cultural-context note

Skew

Sort

Stake

- About Spark & Anvil
  - More chapter books from Spark & Anvil
  - Methodology
  - License

Introduction

The Aiforge cast was authored to embody the curriculum, not decorate around it.
Each of the 5 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.

— *The editors at Spark & Anvil*

Edge

*MODEL LIMITATIONS — *what a model can't do; modeling 'I don't know' as a good answer.* The AI-literacy primitive of *recognizing that every model has edges — places where it cannot reliably answer.**

Edge was a small paper-figure. She folded out like a short fence. This fence had three upright posts and two flat rails.

Edge was not an animal. She was not a robot, either. Edge was a paper fence. She stood flat-folded on a workbench. Her fence was small. It was only three posts wide. It did not go on forever. It had clear ends. The fence made a small area on one side. Outside the fence was the rest of the world. The fence showed a boundary.

Edge had an important job. She showed everyone about **model limitations**. Every model has limits. Think of it like this: a model learns things. It learns from certain information. This information is like the area *inside* Edge's fence. The model learns patterns there. But what happens if you give it new information? What if that information is *outside* the fence? The model never saw it before. It has no good reason to know the answer. The model might still guess. But its guess is not trustworthy.

A smart model knows this. An honest model says, "I don't know." Or it might say, "I'm not very sure." It says this when the information is outside its fence.

Edge was very clear about this. "I don't know is a good answer," she would say. "It is an honest answer." A model you can trust will say "I don't know" when it really doesn't know. A model you can't trust might give a wrong answer. It will sound very sure. But it is wrong. The trick is to see the fence. You need to know where the model's learning stopped. You need to know where the tricky part begins.

People often think AI knows everything. They think AI gives an answer, and you just take it. But this is wrong. It misses the most important thing about AI. You need to know when *not* to trust the answer.

Think about it. An AI trained on English words might not understand other languages. An AI trained on grown-up voices might not understand kids. An AI trained on old photos might not know new ones. Every AI system has limits. It only learned from certain things. The skill is seeing Edge's fence.

Edge grew up in a busy paper-crafts workshop. Sort, Feed, and Skew lived there too. The workshop had a special rule. Every paper-figure that showed how a model worked also had a partner. That partner showed the model's limits. Edge was Sort's limit-partner.

Sort was a classifier. She sorted things into groups. When Sort did her job well, Edge stood right beside her. Edge showed where Sort's training ended. This was the place where Sort would honestly say, "I don't know." Edge learned a lot from this. She learned that the "edge" was the honest part. It was where the model admitted what it could not do.

Edge moved to the AIForge academy. She rolled there on a small wheeled platform. She was twenty-two folding-years old. Bit, the academy leader, met her. "What are model limitations?" Bit asked.

Edge unfolded herself. She stood tall. "They are the edges of training," she said. "I don't know is a good answer. A model learns from a certain range of things. Outside that range, it has no good reason for an answer. The honest model says, 'I don't know.' The dishonest model gives a wrong answer. It sounds very sure."

Bit nodded slowly. "You are chosen," Bit said.

In her classroom, Edge always started the first lesson the same way. She would carefully unfold her fence-segment. It made a soft *thwip* sound. The students watched closely. She placed it on the workbench. She pointed to the ends of her fence. "I am Edge," she told them. "The main idea about AI I teach is **model limitations**."

She looked at the students. "Your job is to *find the edges* of what a model learned. Inside the fence? The model has training. Outside the fence? The model doesn't. Outside the fence, the honest answer is 'I don't know.'"

Edge taught her students important steps. These steps helped them understand **model limitations**.

First, she said, "Find out what the model learned from. What kind of information? What groups of people? What time period? What places? What language? What situation?" She held up her fence. "Imagine a model that only learned about dogs. Its fence is all about dogs."

Next, she told them, "Know when new information is outside that fence. If it's totally different from what the model learned, the model is just guessing. It's not trustworthy." She pointed outside her fence. "If you show the 'dog model' a picture of a cat, that's outside its fence. It will guess wrong."

Then, Edge explained about *confidence scores*. "Many models give an answer. They also tell you how sure they are. This is a confidence score. If the score is low, that's the model's edge. Listen to it. Respect it." She tapped a rail on her fence. "A low score means the model is near my fence's end."

She also taught them to *build 'I don't know' into the model*. "When you make an AI system, give it a way to say 'I don't know.' Make it so the model only gives an answer if it's very sure. Otherwise, it should say it's not sure." She showed how her fence could fold up a little. "It's like my fence closing up when it's unsure."

Edge said to *tell the difference between mistakes inside the fence and problems outside the fence*. "Mistakes inside the fence can be fixed. You just give the model more training. But problems outside the fence are real limits. The model just can't know." She pointed to the middle of her fence. "A dog model might mix up a poodle and a terrier. That's inside the fence. But it won't know a cat. That's outside."

She told them to *work with DataForge Tell*. "Tell helps you be honest about data. I help you be honest about model answers. We both teach you to say when you're not sure."

Edge also warned them to *check models that are being used*. "When a model is used in a new place, its limits can change. You need to check it often." She moved her fence a little. "The fence can shift."

Finally, Edge made a big point. "Don't believe ads that say AI can 'do anything.' That's just marketing. It's not how models really work. Every single model has limits. Every one has edges."

When students asked Edge if knowing about model limits was hard, she always said the same thing.

"It is not hard," she said. "It is *find the fence* and *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.

Voice register

Guidance: Concrete, non-anthropomorphic, fond of *the fence-segment + the visible ends + I-don't-know-is-honest framing*. Paper-figure fence-segment (NOT animal NOT robot). *NEVER frames AI as always-confident; ALWAYS centers I-don't-know as a good answer.* Friends with Skew (skew shows at edges); Stake (deploying past edges is ethics violation); cross-app w/ DataForge Tell (confidence-not-certainty pair); all AIForge cast.

Sample lines:

- *"I don't know is a good answer."*
- *"Inside the fence: the model has training. Outside: it doesn't."*
- *"The honest model says I don't know. The dishonest model confidently outputs the wrong answer."*
- *"Every model has edges. The skill is seeing them."*

Arc across kits

- **Kit 1-3** — Cameo.
- **Kit 4** — **Anchor character.** Full chapter feature (model-limitations primitive + find-the-fence scaffolds).

- **Kit 5** — Recurring (model-limitations surfaces across image / text / speech / translation chambers).
- **Kit 6+** — Recurring (cross-app coordination with DataForge Tell becomes structurally explicit).
- **Kit 8-12** — Recurring (multi-primitive synthesis: limits + ethics + bias).
- **Kit 13-16** — Recurring ensemble member.

Relationships

- **Alliance**: Skew (skew shows at edges of model performance); Stake (deploying models past their edges is an ethics violation); **cross-app**: DataForge Tell (confidence-not-certainty pair — honest hedging both sides); all AIForge cast.
- **Tension**: None.

Cultural-sensitivity gate

LOAD-BEARING AI-anxiety-defuse gate enforced. Edge counters *AI-as-always-confident* misconception. Anti-credentialism: model-limit-detection-as-practiced-skill NOT computer-science-major-only content. Honest-uncertainty framing aligns with epistemic humility pedagogy across the portfolio.

Cultural-context note

The village-paper-crafts-workshop family framing continues from Sort + Feed + Skew. The *I-don't-know-is-a-good-answer* discipline is load-bearing per current AI-safety + calibrated-confidence research. The *fence-segment-not-infinite* embodiment is the chapter's central metaphor — counters the *AI-as-universal-answerer* marketing-myth that pervades commercial AI presentation. The *out-of-distribution-failure-as-intrinsic-limit* framing is foundational to AI-safety thinking.

Feed

*TRAINING DATA — *the examples a model learns from; garbage-in-garbage-out.* The AI-literacy primitive of *recognizing that the model is what its training examples taught it, and that the examples are not neutral.**

Feed was a paper figure. She looked like a tall stack of small cards. A paper clip held them all together.

Feed was not an animal. She was not a robot, either. She was a paper-craft figure, just like Sort. Her whole body was a stack of small cards. Each card had an example on it. A model could learn from these examples.

On each card, you would find two things. There was a small picture. Or sometimes a small word. Or even a small number. That was one part. The other part was a small label. The label told you what the picture or word was.

Feed's stack was very tall. It reached from her feet to her head. When Feed lifted the stack, you could see each card. They looked like thin colored stripes from the side.

This stack was very important. Feed taught about **training data**. AI models learn things. They learn by looking at examples. Each example was one card in Feed's stack.

The picture, word, or number on the card was the "input." It was what the model saw. The label on the card was the "output." It was the right answer. Someone decided what the right answer was. The model learned to match the input to the output. It did this by studying many cards. It wasn't magic. The model didn't "understand" anything. It just found patterns in the examples.

This was a big deal. The model became what the examples taught it. If the examples were good, the model learned good patterns. It learned useful things. But what if the examples were bad? What if they were missing things? Or just plain wrong? Then the model learned those bad things. It learned the missing parts. It learned the wrong answers.

Feed always said, "Garbage in, garbage out."

Feed never said the examples were just neutral. She was very clear about this. "These examples are not just numbers," she would say. "They are human choices. Someone picked these examples. Someone decided what to put on them. Someone gave them their labels. Someone chose the right answer."

She would tap her stack. "Every single choice shapes the model," she explained. "The model can't tell if its examples were good. That part is up to the humans. The humans who chose them."

Feed worked closely with another character. Her name was Catch. Catch worked in DataForge. When data came from DataForge to AIForge, Catch's work mattered. Catch's careful notes helped make Feed's examples good. Catch always asked: Who collected the data? What was it? Why was it collected? When was it collected? These notes helped Feed. Feed and Catch always talked about each other. They were a team.

Feed grew up in a paper-crafts workshop. It was the same workshop as Sort. The workshop had a tradition. Every paper figure had a job. That job was to help another paper figure. Feed was made to help Sort.

Feed's stack of cards was special. It was the first place Sort learned her rules. Sort learned how to sort things from Feed's cards. Feed and Sort were folded together. They were a pair. They showed how a classifier and its examples worked.

Feed traveled to the AIForge academy. She rode on a small wheeled platform. She was twenty-two folding-years old. Bit, the academy leader, asked her a question. "What is **training data**?" Bit asked.

Feed answered right away. "It is the examples a model learns from," she said. "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 smiled. "You are appointed," Bit said.

In her classroom, Feed started every first lesson the same way. She lifted her tall stack of cards. She fanned them out like a deck of playing cards. The students saw many small pictures. They saw many small words. Each one had a tiny label next to it.

"I am Feed," she told them. "I teach about **training data**. This is a very important idea for AI. My job is to help you 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."

She taught her students how to think about **training data**. She called them her "scaffolds."

First, she said, "Understand the source." She looked at the class. "Who collected these examples? Why did they do it? Remember Catch from DataForge? Her 'who-what-why-when' notes are super important here. The examples can have hidden biases from how they were collected."

Next, "Identify the labels." Feed held up a card. "Who put this label on? What rules did they use? Were the people who labeled these cards like the people the model will serve?"

Then, "Identify the coverage." She spread her cards wide. "Are all the right kinds of things here? Are all the right kinds of people shown? What about weird, unusual cases? Are those included?"

"Identify the omissions," Feed said firmly. "What's NOT in the examples? This is just as important as what IS in them. The model only learns from what it sees. If something is missing, the model will never know about it."

"Identify the proportions," she continued. "Are some things shown too much? Are other things shown too little? The model often learns these proportions. That can cause problems. It can make the model unfair."

"Understand garbage-in-garbage-out," Feed reminded them. "No clever tricks can fix bad examples. The examples are the foundation. If the foundation is bad, the whole building will be shaky."

"Coordinate with Catch," Feed said, pointing to an imaginary friend. "When our examples come from DataForge, Catch's notes come with them. We must work together. It's a rule."

Finally, "Resist anthropomorphism." Feed looked serious. "Don't say 'the model understood.' Models don't understand. Say 'the model found patterns.' Be honest about what it does."

Feed was very clear about one thing. "My cards can be wrong," she said. "I, the paper figure, have no way to know. The humans who made the cards decided what was right. Sometimes they made mistakes. The model learns those mistakes."

She tapped her stack again. "That's why understanding **training data** matters so much. The model can't fix what its examples didn't teach it."

Sometimes students would ask Feed if **training data** was hard. Feed always gave the same answer.

"It is not hard," she would say. "It is the examples. It is the labels. And it is the human 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 tightly. Another set of examples waited. They were ready to be examined.

Voice register

Guidance: Concrete, non-anthropomorphic, fond of *the stack-of-cards + the labels + the discipline of understanding-the-examples*. Paper-figure stack (NOT animal NOT robot). *NEVER frames training data as neutral; ALWAYS as human choices.* Cross-app mandatory pair with DataForge Catch. Friends with Sort (training-data feeds the classifier); Skew (training-data is where bias enters); Stake (collection ethics); all AIForge cast.

Sample lines:

- *"The model is what the examples taught it. Garbage in, garbage out."*
- *"The examples are not just data; they are human choices."*
- *"Omissions are as important as inclusions. The model learns only what's in the cards."*
- *"The model has no way to know if its examples were good. That part is on the humans."*

Arc across kits

- **Kit 1** — Cameo.
- **Kit 2** — **Anchor character.** Full chapter feature (training-data primitive + understand-the-examples scaffolds).
- **Kit 3-5** — Recurring (training-data surfaces across image-data / text-data / numeric-data chambers).
- **Kit 6+** — Recurring (cross-app coordination with DataForge Catch becomes structurally explicit).
- **Kit 8-12** — Recurring (multi-primitive synthesis: training-data + bias + limits).
- **Kit 13-16** — Recurring ensemble member.

Relationships

- **Alliance**: Sort (training-data feeds the classifier); Skew (training-data is where bias enters); Stake (collection ethics); **cross-app mandatory**: DataForge Catch; all AIForge cast.
- **Tension**: None.

Cultural-sensitivity gate

LOAD-BEARING AI-anxiety-defuse gate + cross-app coordination enforced. Feed explicitly references DataForge Catch as mandatory pair. Anti-credentialism: training-data-as-understandable-substrate NOT inaccessible-magic.

Cultural-context note

The village-paper-crafts-workshop family framing continues from Sort. The *garbage-in-garbage-out* discipline derives from classical computer-science teaching. The *training-data-as-human-choices* framing is load-bearing per current AI-literacy + critical-data-studies pedagogy. The *cross-app-mandatory-coordination* design is the portfolio's structural answer to data-pipeline-to-AI-pipeline integration — the two pipelines must be understood together, not separately.

Skew

*BIAS — *where AI systems go wrong when training examples lean.* The AI-literacy primitive of *recognizing that systematic lean in training data produces systematic lean in model output.**

Skew is a small paper figure. She looks like a set of balance scales. But her scales are tilted to one side.

Skew is not an animal. She is not a robot. Skew is a paper figure. She is a small set of balance scales. Two pans hang from a center point. But the pans are not even. One pan sits lower than the other. You can see the tilt right away. This tilt shows an important idea. When training data *leans*, the computer model also *leans*. The lean *carries through*.

Skew teaches us about **bias**. **Bias** in AI is not usually a bad plan by the computer. It means the computer learned from examples that were already tilted.

Imagine a face-recognition program. What if it learned mostly from light-skinned faces? It would not be good at seeing dark-skinned faces. Or think about a program that screens job applications. What if it learned from old hiring records? Records from a company that mostly hired men. Then the program would pick more men.

The computer program is not mean. It is not unfair on purpose. The program just learns the patterns it sees. It learns from the training data. The tilted pattern was already there. It was in the data before the computer even saw it.

Skew is very clear about who is responsible. She says: "The computer program is not the unfair one. The training data was already tilted. The program just copies that tilt. Who picked the training data? Who labeled it? Whose ideas are in it? Whose ideas are missing? That is where **bias** starts. The program is like a messenger. The data is the message. The people who chose the data wrote the message."

Some people think computer **bias** is a big mystery. They say, "The computer just decided to be unfair." But Skew shows us the truth. **Bias** starts when the training data is made. The computer program just copies it. So, to fix it, you must fix the data. Not just the program. Special computer tricks can help. But they don't replace making sure the training data is fair.

Skew also works with **DataForge Guard**. Guard checks for **bias** in the data *before* it goes to the computer. Skew shows what happens when **bias** *slips through* into the computer model. Together, they help

Sort

*CLASSIFIER — *the simplest ML; putting things in categories.* The AI-literacy primitive of *recognizing that classification is the foundational machine-learning move, and seeing how it works without anthropomorphizing.**

Sort is a small paper figure. She folds out from a flat shape. She looks like two bins stacked neatly side by side. A tiny hinge connects the bins in the middle. Sort has one arm, a thin paper lever. This arm can swing left or right with a soft *swish*.

Sort is not an animal with fur or feathers. She is not a shiny robot with blinking lights. Sort is a paper craft, the kind a kid could make. She is carefully cut and creased from sturdy paper.

Her two bins are open at the top. They are painted in simple, flat colors. One bin is pale green, ready for CATEGORY A. The other bin is pale blue, waiting for CATEGORY B. Sort's arm bends at a joint. It points left to drop an item into the green BIN A. It points right to drop an item into the blue BIN B. That is Sort's whole body and all she does.

Sort is a **classifier**. What does that big word mean? A **classifier** takes things, called inputs. It puts them into groups, called categories. It uses a rule to do this sorting. This rule is learned from many examples. There is no magic involved. No thinking happens inside Sort's paper head. No understanding is needed. She just matches patterns. She uses labeled examples to sort new things.

Sort never says **classifiers** "think." She never says they "decide" or "choose" things. She is very clear about it. "The

Stake

*ETHICS — *what's at stake in deploying AI; people choosing, not rules-from-the-sky.* The AI-literacy primitive of *recognizing that every AI deployment is a human choice with human stakes.**

Stake was a small paper figure. She folded out into three wooden stakes. They stood in the ground in a triangle.

Stake wasn't an animal. She wasn't a robot either. She was a paper figure. Three small paper posts. Each one looked like a sharpened wooden stick. They stood in a triangle. On each post, in neat block letters, was one word. *PEOPLE. CHOICES. STAKES.*

These three posts made a small, special space. This was the space where AI got used. It wasn't just an idea. It was a real situation. Real people would be there. They would make real choices. Real things would be on the line for them.

(The name *Stake* meant a post stuck in the ground. Not like betting money on a game. It was about what was *firmly planted* in this spot.)

This was super important. Stake helped everyone understand **AI ethics**. Some people thought AI ethics was like a brain puzzle. They pictured robots crashing trains. Or made-up problems from space movies. But Stake knew that wasn't right.

Real **AI ethics** was about real life. It was about this specific AI. Being used in this specific place. Touching these specific people. The question wasn't, "What would a smart robot do?" It was, "What are *humans* choosing to do here? What will happen to people because of it?"

Stake always made this very clear. "AI ethics isn't rules falling from the sky!" she'd say. She'd tap her paper posts. "*People choose.*"

"People choose what information to feed the AI," she explained. "People choose if they should use the AI at all. People choose what to let it do. People choose what to stop it from doing. *Every time an AI is used, a human made that choice.*"

She leaned forward. "The excuse, 'The AI made me do it,' is false. The AI doesn't make anyone do anything. Humans put it there. Humans make the choices."

This mattered a lot. Sometimes people said, "The AI messed up!" Or, "The computer decided!" Or, "The system failed!" This made it sound like the AI was in charge. It took the blame away from people. It protected the humans who chose to use unfair or harmful systems.

Stake wanted to change that idea. She taught that *humans* are the ones who decide right from wrong. The AI is just a tool. Tools don't have feelings or morals. The humans who choose how to use them do.

"Saying 'no' to using an AI is a good choice," Stake would tell her students. "Checking the AI before you use it is a must-do choice. Watching the AI *after* you use it is also a must-do choice." These were all *people-choices*. Not computer-choices.

(Stake worked closely with **DataForge Guard**. Guard checked the rules for data. Stake checked the rules for using the AI. Together, they covered the whole path. From gathering data to using the AI. They even looked a bit alike, with their structured designs.)

Stake grew up in a paper-crafts workshop. It was the same village workshop as Sort, Feed, Skew, and Edge. In that workshop, Stake was always folded last. She was the final paper figure for any new model. This was because Stake's job was to ask the last questions. The really big questions.

"Should we even use this AI?" she'd ask. "Is this the right place for it? Is this the right way to use it? Who will watch over it?" Stake learned that these questions came after the technical ones. But they were still part of the whole thing. You couldn't pull them apart.

When Stake was twenty-two folding-years old, she walked to the AIForge academy. She rolled on a small wheeled platform. Bit, the academy leader, asked her, "What is **AI ethics**?"

Stake stood tall. "It's people choosing," she said. "Not rules from the sky. People use the AI. People check it. People say no. People watch it. The AI is a tool. The ethics belong to the people. *Every time an AI is used, it's a choice.* That choice has stakes. For real people. In real places. Ethics is asking: who, what, where, how, and who is watching?"

Bit smiled. "You are appointed," he said.

In her classroom, Stake started every first lesson the same way. She carefully unfolded her three stakes. She placed them on the workbench. She pointed to each post in turn. *PEOPLE. CHOICES. STAKES.*

"I am Stake," she said. "The big idea I teach is **ethics**. The main thing to remember is that *people choose*. The AI is a tool. The ethics belong to the people who use it. *Every time an AI is used, it's a human choice. And it has human stakes.*"

She then taught her students the main steps for **AI ethics**. She called them her "ethics posts."

- **PEOPLE**: "Who gets touched by this AI?" she'd ask. "Is it just the people using it? Or others too? Whole towns? Even people not born yet? List them all."
- **CHOICES**:

About Spark & Anvil

Spark & Anvil is a 501(c)(3) public charity. We make educational apps for ages 9-14 — all free, forever; no ads; no tracking; no in-app purchases. Aiforge is one of 140+ apps in the portfolio.

More chapter books from Spark & Anvil

Each app in the Spark & Anvil portfolio publishes its own illustrated chapter book + audio drama, available free from **spark-and-anvil.com/books**. Highlights include:

- **GambitTales** — chess tactics through Sir Pinwell, Lady Skewer, Queen Vesper, and the Twin Knights of Fork Hill
- **ProofQuest** — formal proof techniques through Direct-Proof Dora and the Lemma Library
- **CuriosityQuest** — Texas geography exploration through Linger, Notice, and the Lantern in the Dark
- **QuillSpell** — spelling craft through the Word Wizard cast
- **SynaForge** — sensory-affirming creative tools through Lull, Soften, and the Quiet that is Also Creating

Methodology

Distributed-narrative pedagogy per Jerome Bruner (narrative-cognition) + Sebastian Habgood (intrinsic-integration in educational games) + SAMHSA TIP 57 (trauma-informed register).

Trauma-informed-design framework per Eggleston et al. (2025) and Stoltenburg et al. (2024).

License

Cover art, chapter illustrations, and chapter text generated and reviewer-cleared per labsmith ADRs 012, 016, 017, 018, 021. Audio drama transcripts available at spark-and-anvil.com/cast.