Designing Ethical AI for Learners

Generative AI Playbook for
K-12 Education

Quill

Hello Reader!

For the last six years, our team at Quill.org has been building our own AI models to provide students with AI-powered learning tools. Our tools have now helped more than ten million students become stronger readers and writers. **We wrote this playbook so that education leaders around the world can learn from our hard-won insights, using our playbook to both build their own AI learning tools and to ask better questions while evaluating AI from a potential partner.** In this document, we will take you behind the scenes to share our entire process, step by step, for creating our AI-powered tools so that you can understand the key components to building ethical AI that is highly effective for student learning.

Educational institutions are now tasked with addressing the lowest literacy and math scores in the United States in the past 30 years. Low-income, Title I schools—those that were hardest hit during the pandemic—stand to benefit most from AI solutions that can help accelerate learning outcomes. However, there is a real risk that, as organizations begin to scale up their use of AI, poorly designed tools will harm students by creating ineffective, frustrating learning experiences.

That's where Quill comes in. As a 501(c)3 nonprofit, we're focused on serving students whose schools may not have the resources they need to excel, and we've intentionally designed our AI tools with those students in mind. When a student is completing an activity on Quill, they're receiving feedback in real time. To do so, we are responsible for building AI that accurately assesses student work and provides guidance that is meaningful, accessible, and aligned with the coaching they would receive from a veteran educator.

We take this responsibility seriously: over the last six years, we have built more than 300 unique AI models to evaluate and coach student writing. The biggest lesson we've learned? **AI is malleable.** People can build processes and systems to control and improve AI performance. While out-of-the-box AI can be unpredictable, carefully annotated datasets and robust data evaluation infrastructure make it possible to customize AI to effectively and reliably engage students.

This playbook contains the most important takeaways from our efforts, written in collaboration with co-author Afua Bruce, a leader in the responsible AI development movement, who was instrumental in translating our process into actionable insights. We hope that this guidance will help you deploy your own AI-powered tools and ask the right questions when evaluating them.

Sincerely,
Peter Gault, Maheen Sahoo, and Afua Bruce

Quill's Four Steps for Developing Ethical AI for K-12 Education

Below is an overview of the four steps we use to build and evaluate our AI. Later in this playbook, we'll do a deep dive and explain these steps in more detail.

Step 1: **Research Before You Code**

To Shape the Future of AI Learning, First Study What's Effective in the Classroom Today

Before writing a line of code or creating a single AI prompt, we work with researchers and educators to precisely define what successful learning looks like. We collect and analyze a large set of examples from the classroom. Why? Because a crystal-clear definition of success is essential, not only for training the AI, but also for assessing and improving its performance. You won't be able to evaluate your AI's performance until you have defined what effective learning actually looks like and means in your specific context.

Step 2: **Bring Your Own Dataset**

Once You've Defined Effective Learning, Create Your Own Dataset that Demonstrates It

Educators, not the generic training of an off-the-shelf large language model (LLM), should be making judgment calls about a student's work. By providing many examples of educators' real judgments, you can modify an AI's "thinking" process and enable the AI to learn from and replicate an educator's judgment calls. In our work, we build a dataset of 50 to 100 examples of authentic student responses paired with high-quality feedback written by educators for every writing task on Quill. These 50-100 examples are provided to the AI every time it evaluates a new student response. Rather than relying on the LLM's default training, the AI draws on the teacher feedback to generate its own coaching for students. By having educators build the training datasets that are embedded into the AI, educators become the creators, rather than the recipients, of AI learning tools.

Step 3: **Evaluate Your AI Early & Often**

You Can't Improve What You Don't Measure

The hardest part of AI development is evaluation - you need to constantly assess the actual student experience and outcomes to ensure that your AI is reliable and equitable. Our ten full-time curriculum developers at Quill—primarily former classroom teachers—manually evaluate over 100,000 student responses each year. This process helps us identify where we're excelling and where we need to improve. Organizations that are serious about building ethical AI must make significant investments in thoroughly evaluating how well their AI operates at scale.

Step 4: **Create Your Own Council**

It's Essential to Collaborate with Educators

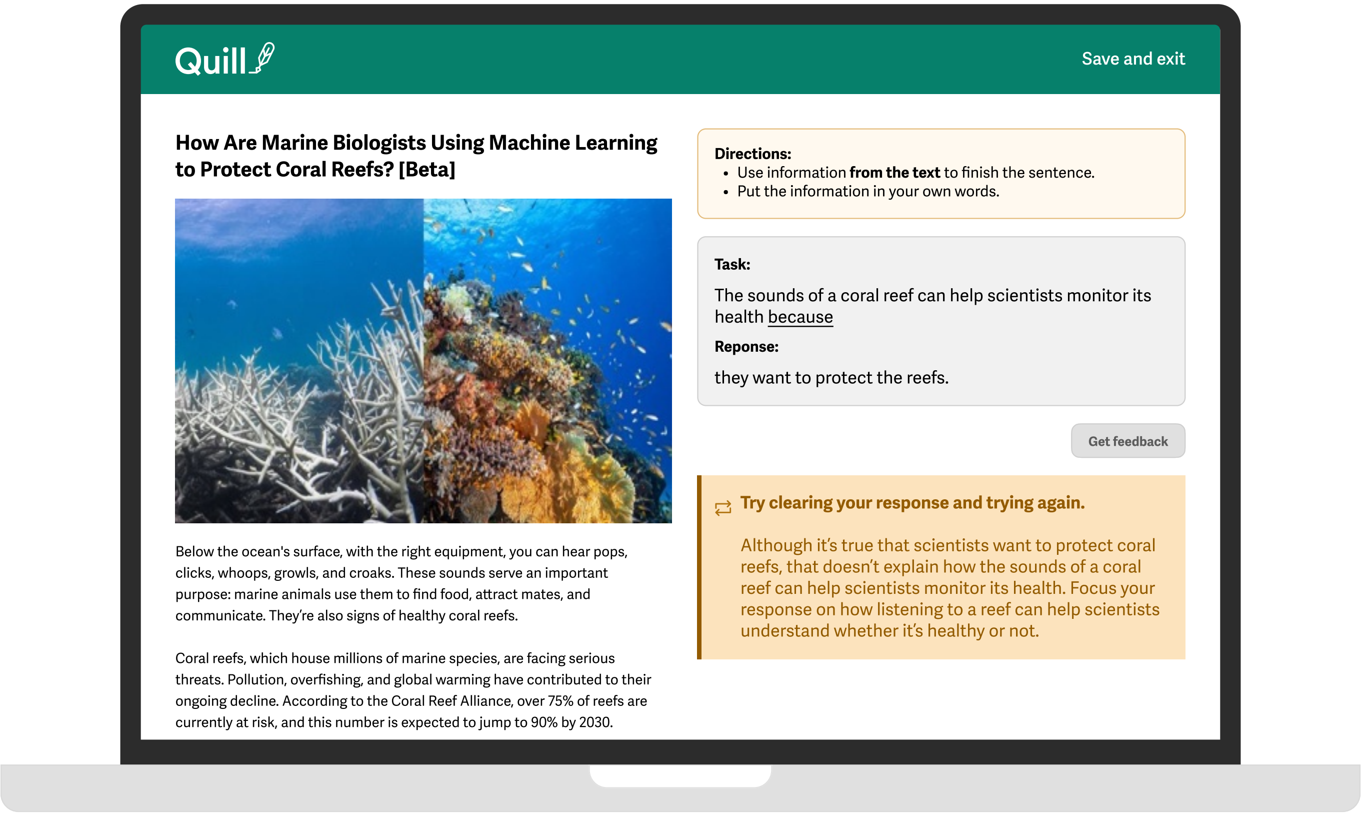
Educators must be partners in development. At Quill, we work with over 300 teachers in our Teacher Advisory Council to evaluate and review every single AI-powered writing prompt we create. They don't just look at the AI once: we complete three rapid-cycle evaluations with our Advisory Council and precisely measure how much the AI improves each time. We don't release our AI to the general public until we are confident that it has reached our key reliability benchmarks.

Meet Quill and The "Why" Behind Our Work

Now more than ever, students need strong critical thinking skills to effectively navigate the world around them and distinguish between fact and fiction. Students must be able to use evidence to construct strong arguments, while also catching their incorrect assumptions. At Quill, we develop writing prompts that help students meaningfully engage with non-fiction, informational texts and foster the skills they need to effectively share their thoughts in writing. While we offer six literacy tools, this playbook focuses on the AI development for Quill Reading for Evidence, our most sophisticated learning tool.

To understand how we build our AI, we first need to show you how students learn through cycles of writing, feedback, and revision. Students begin each activity by reading a short (600–800-word) nonfiction source text. After reading, students use the most relevant information from the text to complete a series of open-ended prompts. Each writing prompt is a claim, and students are asked to use evidence from the source text to support that claim. Once they write and submit their claims, students receive immediate, responsive coaching from our AI. Students can then engage in up to five cycles of feedback and revision; through these cycles, students rapidly strengthen key literacy skills. In a single 15-20 minute session, students practice reading comprehension, cite relevant evidence, paraphrase information from the text, and support claims with logical reasoning.

Throughout this section of the playbook, we will do a deep dive into one Reading for Evidence activity to show you how we created the AI that powers it.



How Are Marine Biologists Using Machine Learning to Protect Coral Reefs? [Beta]

Below the ocean's surface, with the right equipment, you can hear pops, clicks, whoops, growls, and croaks. These sounds serve an important purpose: marine animals use them to find food, attract mates, and communicate. They're also signs of healthy coral reefs.

Coral reefs, which house millions of marine species, are facing serious threats. Pollution, overfishing, and global warming have contributed to their ongoing decline. According to the Coral Reef Alliance, over 75% of reefs are currently at risk, and this number is expected to jump to 90% by 2030.

Directions:
- Use information **from the text** to finish the sentence.
- Put the information in your own words.

Task:

The sounds of a coral reef can help scientists monitor its health because

Reponse:

they want to protect the reefs.

Try clearing your response and trying again.

Although it's true that scientists want to protect coral reefs, that doesn't explain how the sounds of a coral reef can help scientists monitor its health. Focus your response on how listening to a reef can help scientists understand whether it's healthy or not.

The text explains how biologists are using machine learning to protect coral reefs. It comes from our new "Building AI Knowledge" STEM curricular offering, designed to teach middle and high school students about how AI works and its impacts on society. This particular activity, part of a unit on AI and environmental protection, examines how researchers are using audio recordings and AI analysis to monitor the health of coral reefs. The rest of the playbook will walk you through each step of our development process to showcase how we carefully construct this activity's AI feedback loop.

Step 1: **Research Before You Code**

To Shape the Future of AI Learning, First Study What's Effective in the Classroom Today

Before you begin developing AI for education, you must first be able to define exactly what effective learning looks like. That way, you know what you expect the AI to do. We expect our AI at Quill to provide coaching that mirrors a veteran educator's feedback. We must first identify the components of a veteran educator's feedback in order to train AI to meet that expectation. By providing high-quality feedback at the level of veteran educators, teachers can trust that our tools – which are built to support and never replace educators – meet their needs.

To define effective learning, you can start by simply collecting examples of what works and doesn't work—no AI required. At Quill, we start with **paper-based testing:** we build hard-copy exercises and try them with students. In the photo shown here, our Executive Director, Peter Gault, uses dozens of paper strips to test potential pieces of feedback to learn which ones are most useful. Through this extensive testing process, we've identified three core principles for designing effective AI-powered learning experiences: (1) carefully design the student writing prompts to be able to consistently assess responses; (2) clearly define the qualities of exemplary student responses to determine when a student should receive feedback; (3) identify and deliver meaningful guidance in every piece of feedback students receive.

Principle 1

Design an Effective Writing Prompt

One key insight from our testing is that our AI feedback is more effective when the students are all basing their writing on the same source text. When there's no source of truth, students may be completing prompts using prior knowledge, personal experiences, or other sources of information, and it becomes difficult to consistently and reliably assess the strength and accuracy of their writing across thousands of responses. As a result, each Quill prompt is designed from a specific source text. This instructional approach is called **text-dependent prompts**: exemplary student writing must include specific ideas and details from the given text. Not only does this improve AI grading, since the AI can compare directly to the source, but it also delivers a research-validated pedagogical strategy into the classroom. For more than fifty years, researchers have found text-dependent questions to be a "solid foundational practice" (Henry 2022). In the era of Generative AI, this strategy takes on a new dimension for engaging students in deeper and more effective learning and building literacy skills.

Principle 2

Define an Exemplary Student Response

Each time a student writes a response, we hope to see them use specific evidence from the text to support a claim. When students write vague or inaccurate responses, our goal is to help them revise and strengthen their work. In order for the AI to either recognize when a student has written an exemplary response or coach a student to revise their work to get there, we first need to define what an exemplary response is. The writer has their initial idea of the exemplary response before they begin drafting. Once an activity is written, they assess hundreds of examples of student writing to determine the features of a strong response in practice. Where is the boundary between what needs revision and what is acceptable as a final response? By precisely defining these criteria, we can train the AI to look for and coach toward success. In the example below, you can see how a student who initially wrote a vague response completed two rounds of revision to strengthen their writing and clearly explain how scientists use sounds to assess a reef's health.

Prompt:

Use evidence from the text to support this claim:
The sounds of a coral reef can help scientists monitor its health because...

Student Response:

The sounds of a coral reef can help scientists monitor its health because...
they want to protect the reefs.

Feedback:

Although it's true that scientists want to protect coral reefs, that idea doesn't explain how the sounds of a coral reef can help scientists monitor its health. How does listening to a reef help scientists understand whether it's healthy or not?

Student Response:

The sounds of a coral reef can help scientists monitor its health because...
coral reefs make different sounds.

Feedback:

Healthy and unhealthy coral reefs sound different from each other—that's true! Now revise your response to be more specific. What do healthy and unhealthy coral reefs sound like?

Student Response:

The sounds of a coral reef can help scientists monitor its health because...
healthy coral reefs make crackling sounds and unhealthy reefs are quiet.

Feedback:

Nice work! You used information from the text to explain why the sounds of a coral reef can help scientists monitor its health.

Principle 3

Deliver Effective Teacher Feedback:

For students to reach an exemplary response, we need to provide effective feedback and coaching. Extensive research and education scholarship describe principles of effective feedback for students. We started with that research, and then we got even more specific. We asked ourselves, “How does a veteran educator provide guidance to their students?” To answer this question, our team first collects a series of student responses and then asks a diverse cohort of teachers to write out the feedback they would give if they were working with each student individually.

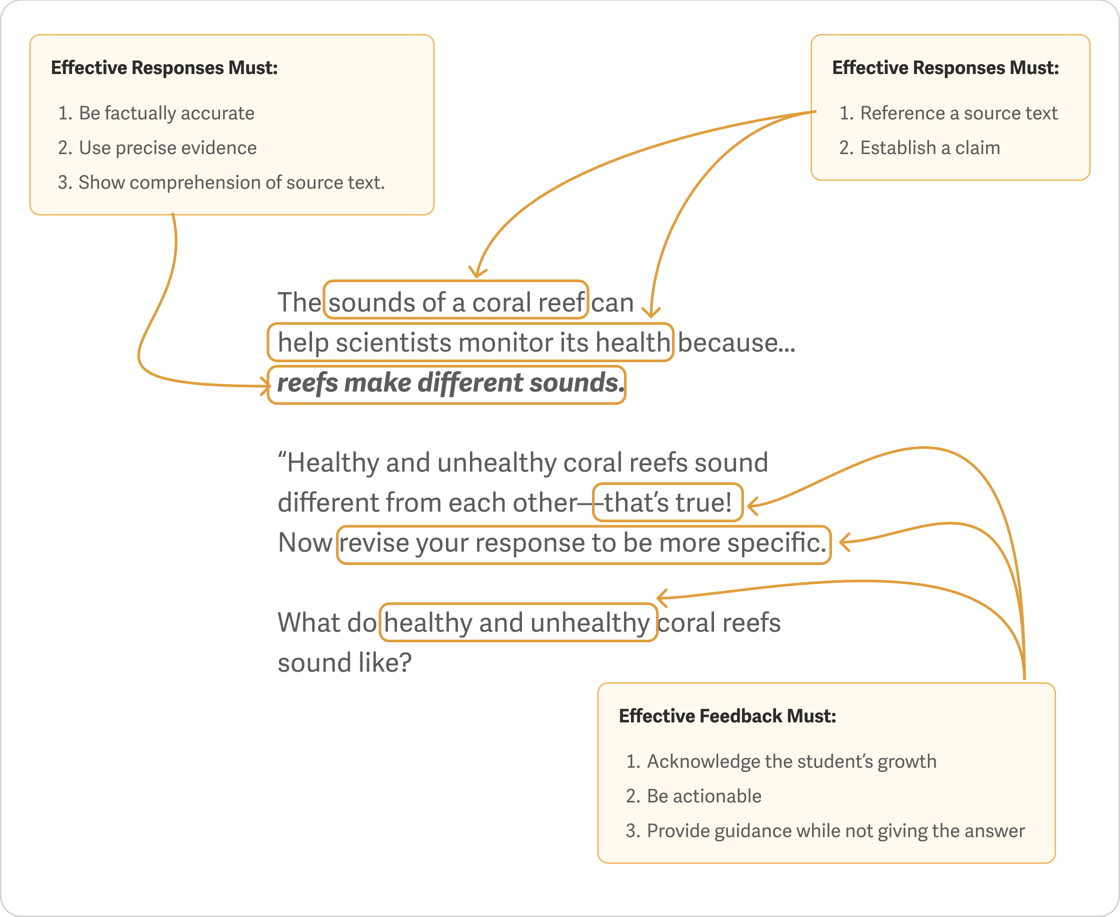
By reviewing hundreds of examples, we find common patterns in how these educators coach student writers. The sample below illustrates how three educators have used different approaches to support the writer. Creating a library of examples like these can provide a valuable reference point for developing a clear and consistent definition of what effective feedback looks like for your organization.

Writing Sample:

The sounds of a coral reef can help scientists monitor its health because…
it is difficult to monitor the sounds of the ocean.

Educator #1	Educator #2	Educator #3
”Try again! Write about a **reason** that the sounds of a reef are helpful.”	“You’re right that the text says that the process is difficult, but that’s not information you need to focus on here. For this sentence, you need to find an explanation for why the process works. Try using ‘because’ to explain how listening to a reef can help scientists understand whether or not it’s healthy.”	“Let’s think about this another way: How do the sounds of a reef give scientists clues about its health?

Returning to our activity on protecting coral reefs, you can see how these principles are represented within a single writing task. The prompt, feedback, and grading are carefully designed based on these principles for effective learning.



Effective Responses Must:

1. Be factually accurate
2. Use precise evidence
3. Show comprehension of source text.

Effective Responses Must:

1. Reference a source text
2. Establish a claim

The sounds of a coral reef can help scientists monitor its health because... ***reefs make different sounds.***

"Healthy and unhealthy coral reefs sound different from each other—that's true! Now revise your response to be more specific.

What do healthy and unhealthy coral reefs sound like?

Effective Feedback Must:

1. Acknowledge the student's growth
2. Be actionable
3. Provide guidance while not giving the answer

Well-defined principles provide clear guidelines for effectively training the AI to meet your goals. Without investing the time to establish these in advance, you risk unreliable results with different outcomes and varying degrees of success each time.

Step 2: **Bring Your Own Dataset**

Once You've Defined Good Learning, Create Your Own Dataset that Demonstrates It

Once you have defined your guiding principles, you're better positioned to build AI that effectively works toward them. With Generative AI, a student or teacher often provides input to the AI, such as the written response or the requirements for a lesson plan. Based on this input, the AI generates an output, such as a piece of feedback or a lesson plan.

When you ask the AI to provide its output, you are relying on the underlying training of the large language model (LLM) to generate a helpful response. However, rather than relying on the underlying training data, you can shape the output yourself by providing the AI with a **custom dataset of examples** to guide its output. At Quill, we do not simply send the student's response to an LLM. Instead, for every student writing prompt developed, the LLM is first provided with **50-100 sample student responses** (anonymized and aggregated) paired with **real teacher feedback** written by our team. Through these additional inputs (sample student responses) plus outputs (teacher feedback for those responses), we teach the AI how to assess a student's response. In total, we send **about 5,000 to 8,000 words** of directions, sample student responses, and exemplary teacher feedback to the LLM for every student response we ask the AI to evaluate. Here are a few examples of the student responses and teacher feedback we send to the LLM:

Feedback Topic: Student response must be specific and detailed	
Student example:	**Teacher feedback:**
The sounds of a coral reef can help scientists monitor its health because ***healthy and unhealthy reefs sound different.***	"That's true! Now make your response more specific. How do healthy reefs sound compared to unhealthy reefs?"

Feedback Topic: Student response must be based on evidence from the source text	
Student example:	**Teacher feedback:**
The sounds of a coral reef can help scientists monitor its health because ***unhealthy reefs have more high-frequency sounds.***	"Try clearing your response and starting again. It's true that unhealthy reefs give off more high-frequency sounds, but that idea isn't in the text."

Feedback Topic: Student response must be a factual statement, not an opinion

Student example:	**Teacher feedback:**
The sounds of a coral reef can help scientists monitor its health because ***protecting coral reefs is the most important conservation effort.***	"Use an idea from the text instead of giving an opinion. If the opinion you wrote is expressed in the text, add who believes it to your response."

To develop this dataset, our team works with our Teacher Advisory Council, a group of more than 300 teachers across 42 different states—over 68% of whom work in Title I schools—to collect hundreds of responses. Then, we write our own student-friendly feedback to complete each dataset.

Once we have built a set of student responses and feedback, there are two primary AI techniques for embedding this dataset into the AI: **multi-shot prompting** and **finetuning**. With multi-shot prompting, you simply include the examples in your request to the Generative AI model. You can write, "I am going to provide five examples to show you how to assess a student's writing." In AI parlance, this is called "five-shot prompting." A shot is an example provided to the LLM. With finetuning, you typically upload a large dataset in advance, and the AI model learns from those examples. These two techniques can be used independently or together to customize and shape the AI's output.

At Quill, we have used both techniques: we have created fine-tuned models by training the AI on datasets of thousands of student responses paired with teacher feedback, and, when using multi-shot prompting, we provide 50-100 student responses to a Large Language Model. Both techniques create a **"thick wrapper"** around the AI—this is the key differentiator between Quill and **"thin wrapper"** AI tutoring tools. With a "thin wrapper" tool, a student's response is sent straight to the LLM without any specialized training or datasets other than those used to build the LLM itself. The problem with thin wrapper tools is that they are, by definition, not customized to the needs of students or teachers. As a result, responses will solely depend on the large language model's default training data.

For example, generative AI systems are often trained to be as helpful as possible to the user. For educational tools, this can create problems. To be helpful, the AI tends to reveal the answer to the student instead of enabling the student to figure it out themselves. This means that learners miss the opportunity to think critically and construct their own answer, doing the "heavy lifting" themselves. Some AI developers try to address this by including directions in the prompt, such as, "Don't tell the answer to the student." However, at the time of this playbook's publication, our research has found that LLMs don't respond well to simply being told what not to do. Instead, you need to show the LLM what to say instead; this gives the AI a model for how to guide the student without revealing the answer.

By feeding the LLM a variety of student responses paired with teacher feedback, we ensure that the AI feedback is tailored to be highly effective in bespoke contexts. This is a **quality-over-quantity** approach. By slowing down and building custom datasets to “get it right,” we build our capacity to give highly accurate and effective feedback, rather than trying to do too many things at once and providing ineffective learning experiences.

Step 3: **Evaluate Your AI Early & Often**

You Can't Improve What You Don't Measure

The hardest part of responsibly and ethically developing AI is evaluating whether a tool is effectively engaging learners from many different backgrounds. At Quill, having each of our ten full-time curriculum developers manually evaluate more than 100,000 student responses each year to ensure that our AI is reliably evaluating student writing and providing appropriate feedback is a process that takes thousands of hours of labor. But this work enables us to recognize where we are excelling and where we need to improve.

To evaluate our AI, we first create a **benchmark evaluation dataset** and then run a series of **A/B tests** to determine which training data is needed to meet our benchmark evaluation standards. Similar to the way we build the LLM prompt's dataset, we develop this dataset by collecting and manually grading at least 300 responses to serve as our benchmark evaluation dataset. It's critical that the benchmark dataset responses are distinct from the training dataset examples provided to the LLM prompt through shots and finetuning. If you test the AI with the same examples you used to train it, it isn't a test at all: you've already provided the AI with every answer. This makes it more difficult to accurately predict how the AI will perform when it encounters student responses it hasn't seen before.

Every innovative AI company builds its own benchmark datasets, and AI companies routinely update their datasets as their AI becomes more sophisticated. For example, in 2024, Google's DeepMind team realized that the questions in existing fact-checking datasets were too easy to answer and didn't reflect the complexity of the questions that AI systems received in the real world. As a result, they introduced a new benchmark dataset called LongFact. With 16,000 complex questions and answers, LongFact made it possible for the team to better assess how well different LLMs performed these tasks; this, in turn, helped them better identify areas to improve. Any organization that isn't building its own benchmark data has no means of evaluating its AI performance.

At Quill, after building our 300-response benchmark evaluation dataset, we send these 300 responses to the AI so that it can grade the responses and generate feedback. We then compare the human's grading and feedback with the AI's. In the initial trials, the AI's grading aligns with our grading some of the time, and the gaps in performance highlight exactly where we need to revise. For example, an initial custom training dataset may contain 50 excellent examples of teacher-written feedback, but we don't know if we selected the most useful and representative 50 examples of student misconceptions that the feedback should address until we test with unique responses. Based on the results of the initial trials, we create a new version of the training dataset with additional examples of authentic student writing paired with handwritten teacher feedback. We feed that into our Generative AI model to see if this new dataset leads to improved grading accuracy.

This process is powered by our A/B testing infrastructure. Our first dataset is Trial A, while our second, improved dataset is Trial B, and by comparing the two side-by-side, we see if the new dataset leads the AI to perform better than the previous one. Although it's called A/B testing, as if there are only two trials, in practice, we do far more than two trials. We may revise our prompt examples and evaluation dataset through dozens of trials before we reach our goal and are ready to move an activity into the next stage of our process.

The A/B testing process is by far the most important part of this playbook. It is the iteration cycle that gets us from an AI feedback loop that works some of the time to an AI feedback loop that is consistent and reliable in evaluating student writing. Thanks to our A/B testing infrastructure, we can quickly test multiple competing approaches side-by-side to determine the best custom dataset for a particular activity. In addition to testing new datasets, we also embed custom directions and guidelines into the AI to enable better student feedback.

In the example below, you can see how the AI's evaluation in Test B outperformed Test A.
In **Test A**, we highlighted three examples of the AI marking student responses as weak responses that require revision, where our team of curriculum developers determined that these same responses were all strong, exemplary responses. In **Test B**, which used the updated dataset, the AI correctly identified that these were all exemplary responses.

Sample Student Response	Gen AI: Test A	Gen AI: Test B
The sounds of a coral reef can help scientists monitor its health because ***healthy reefs make loud noises and unhealthy reefs are quiet.***	That's a good start! Now be more specific. What do healthy coral reefs sound like?	Nice work! You used information from the text to explain why the sounds of a coral reef can help scientists monitor its health.
The sounds of a coral reef can help scientists monitor its health because ***all the animals living in a healthy reef make them noisier than unhealthy reefs.***	It's true that healthy reefs and unhealthy reefs sound different. Now add to your response. How can scientists use their noises to monitor their health?	Nice work! You used information from the text to explain why the sounds of a coral reef can help scientists monitor its health.
The sounds of a coral reef can help scientists monitor its health because ***unhealthy reefs are quiet, lacking any animal noises.***	Unhealthy reefs are quiet–that's true. But it doesn't explain how the sounds of reefs help scientists. How do the sounds of a reef help scientists monitor a reef's health?	Nice work! You used information from the text to explain why the sounds of a coral reef can help scientists monitor its health.

Step 4: **Create Your Own Council**

It's Essential to Collaborate with Educators

AI is never perfect in its first version—collaboration and iteration are required to take an AI tool from good to great. An effective AI tool needs authentic inputs from the people who use it; until that data is present, it will never truly be representative. Before we release a new activity to a wider audience, we share it with our 300-member Teacher Advisory Council for **three distinct rounds** of testing and feedback. Between each of these rounds, our team evaluates the AI's performance and iterates based on these findings.

To begin, in **round one**, the activity is shared with our Teacher Advisory Council, where teachers are invited to use the activity with their students. This is typically the first time the activity is used in a classroom scenario, and we have the chance to see the AI feedback in action. After receiving at least one hundred sessions, our team analyzes the feedback. We look at various metrics: What percentage of students wrote an exemplary sentence on their first try? How many reached an exemplary response by their fifth attempt? We then make adjustments and improvements to our training dataset. Then, we repeat the entire process in **round two** and **round three** when we open the activity to broader groups of teacher users to receive additional activity plays. We want to ensure that we've adequately considered the diversity of student responses before we are, at last, ready to publish the activity to all Quill teachers.

Carmen Adamucci
Saint Monica Preparatory
Saint Monica Preparatory

Juan G. Alvarado
Valley View High School
Hidalgo, TX

Alesha Cary
Madison-Ridgeland Academy
Madison, MS

Rebecca Foland
Waukee Public Schools
Waukee, IA

Audrey Gebber
Gulf Coast High School
Naples, FL

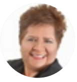
Deana M. Harris
Thorndale High School
Thorndale, TX

Jasmine Hobson Rodriguez
Hesperia High School
Hesperia, CA

Jennifer James
Vinemont High School
Vinemont, AL

Meleighsa McLaughlin
James Clemens High School
Madison, AL

Sera Ramirez
Fort Stockton High School
Fort Stockton, TX


Elma Ruiz
Port Isabel Junior High
Port Isabel, TX

Elizabeth Tanner
Westwood High School
Mesa, AZ

Here are 12 of the 300 educators on our advisory panel. We're grateful that they have dedicated their time to strengthening our AI so that it can better serve millions of students. By working with our Teacher Advisory Council, we are able to ensure that our AI models and texts are not only as effective as they can be, but also well-grounded in (and responsive to) the realities of the classroom.

Conclusion

Asking the Right Questions

Our approach to AI development prioritizes quality, customization, and continuous evaluation. By crafting a custom dataset for each writing prompt and evaluating our performance against that dataset, we can train an AI system to reliably and effectively provide feedback to students. As you develop your own AI-powered learning tools or evaluate products built by others, here are the important questions to ask to assess how effectively and ethically the AI will engage students:

Question 1

Is the AI's approach based on proven teaching methods or research on how students learn, rather than general-purpose AI repackaged for schools?

Look for tools custom-built for particular learning goals rather than a generic AI that hasn't been tailored to classroom needs.

Question 2

What data did the AI learn from and does the AI have a clear idea of what a good answer looks like for each writing prompt?

Look for AI tools that leverage thousands of real student samples to provide custom training to their AI, ensuring that the AI's grading is accurate and relevant for the specific task at hand.

Question 3

What is the composition of the team providing custom training to the AI?

Look for organizations that employ full-time former educators to train their AI models in partnership with their technical and product teams.

Question 4

Does the tool give students specific feedback and suggestions to improve their writing, instead of generic feedback that could apply to any answer?

Look for tools that provide students with very specific feedback that is grounded in classroom learning, rather than AI tools that provide one-size-fits-all comments.

Question 5

What systems do you have in place to ensure the AI's feedback is accurate and helpful?

Look for organizations that deploy evaluation systems to constantly monitor the AI's feedback, alerting the team to issues so they can continually retrain the AI if needed.

Question 6

Were teachers involved in developing and fine-tuning this AI tool to ensure its effectiveness in actual classrooms?

Look for organizations that work with educators at every stage of the process to continually evaluate and improve the AI.

All these questions guide Quill's AI development. As we continue to develop and build our suite of generative AI tools, we aim to set a standard for the responsible and ethical use of AI in education, ensuring that AI technology serves as a bridge, rather than a barrier, to strong student learning outcomes.

Reach the authors at **EthicalAI@quill.org**
to share your thinking on Ethical AI

Quill