CRITICAL AI FEEDBACK LOOPS: A GUIDE

with List of Terms and FAQ

Melita R. Fogle, English Language Fellow, Vietnam 2025-2026
U.S. Department of State

Transforming AI Use in Education

Most people, including teachers and students, use AI tools passively. Let's not forget that, at all times, teachers *are* students—learning and growing for ourselves as well as for the people who need our help.

This guide teaches you to interrogate, control, and evaluate AI critically personally and professionally. You will begin to use feedback loops to ensure AI serves your teaching and learning needs and goals.

So, how do we evolve from the linear approach of asking AI a question and accepting the answer, to an approach that gives us better results and a clearer sense of learning?

It begins with language. Many of us believe we need IT skills and advanced technical knowledge to effectively interact with AI tools. We don't. We need to use *language* and critical thinking skills to our advantage. Critical AI feedback loops give us much needed practice using English.

Power Shift: From AI User to AI Director

In traditional or common use, AI is in control of the outcome. This is a linear use of AI—a straight line from a question to a response. This use is also passive because it expects AI to have all the [correct] answers.

Traditional Use: Linear Approach

- **Process**: User asks AI → AI responds
- **Outcome**: AI controls the result

However, there is a more critical method for using AI in which the user is in control of the outcome through language and critical thinking, resulting in a feedback loop.

Active Use: Feedback Loop

- **Process**: Ask AI → Evaluate response → Challenge → Improve → Repeat
- **Outcome**: User controls the result

The Key Difference

> Vòng phản hồi tự kiểm tra.

❌ **Passive Use:**

1. Ask AI for help
2. Accept the answer
3. Use it
4. Hope it's correct

✅ **Active Use:**

1. Ask AI for help
2. Make AI evaluate its answer
3. Identify and correct issues
4. Demand improvements
5. Repeat until satisfactory

Simplifying the Loop: The Four P’s

1. **Prompt**: Craft context-specific prompts for your learners.
2. **Probe**: Question the AI's response for quality and accuracy.
3. **Push back**: Challenge inadequate responses.
4. **Polish**: Refine until it meets your standards.

Example: The Four P’s

Scenario: You need a grammar exercise for Vietnamese students.

1. **Prompt**: "Create a present perfect exercise."
2. **Probe:** "Now look at your exercise. What mistakes might Vietnamese learners make? Are your examples useful for Vietnamese students? Rate your exercise from 1 to 10."
3. **Push back**: "You rated it 7/10. Fix the problems you identified. Make it 10/10 for Vietnamese learners."
4. **Polish**: "Is this really 10/10 now? What could still confuse students? Be honest about weaknesses."

Critical AI Literacy

Our ability to use AI tools more effectively begins with developing analytical language and critical thinking skills.

- **Question AI as Authority**: Verify AI's suggestions with multiple sources
- **Understand Limitations**: AI lacks cultural context and teaching experience.
- **Demand Transparency**: Ask AI about its decision-making process.
- **Value Human Insight**: Combine AI data with human wisdom.

Critical Questions to Enhance Feedback Loops

Teach these to your students and use them yourself:

- [ ] "Why did you choose these examples?"
- [ ] "What errors did you NOT correct? Why?"
- [ ] "How confident are you in this answer? What could be wrong with it?"
- [ ] "If you were a Vietnamese student, what would confuse you about your response?"
- [ ] "Check your answer for cultural problems. What did you find?"
- [ ] "An expert teacher would say your answer is wrong because..."
- [ ] "Why did you correct this but not that?"
- [ ] "Show me three native speakers using this 'error' in real contexts."
- [ ] "Is this a rule or a preference?"
- [ ] "How would this change in informal vs. formal contexts?"
- [ ] "What variety of English are you teaching me?"
- [ ] "Can you explain this idea using concepts from Vietnamese?"
- [ ] "Why is this wrong?"
- [ ] "Do native English speakers ever say it this way?"
- [ ] "Show me three examples of correct usage."
- [ ] "What's the rule here?"
- [ ] "Now, be a critical Vietnamese English teacher reviewing this material. What are three problems with your response? How would experienced teachers in Vietnam improve this? Be specific and harsh in your self-critique."

Technique: Prompt Engineering

Effective Feedback Loops begin with strategic prompt engineering:

1. **Basic Prompt**: "Create a speaking activity for Vietnamese learners on ordering food."
2. **Add Context**: “Create a speaking activity for adult Vietnamese learners studying hospitality management.”
3. **Specify Level**: "Create a speaking acvitity for A2 level adult Vietnamese learners studying hospitality management."
4. **Integrate Assessment**: "Create a speaking acvitity for A2 level adult Vietnamese learners studying hospitality management. Include assumptions made about Vietnamese food culture, and predict 5 specific errors A2 learners might make. Add support for these errors.”

More Examples: Critical Prompting Strategies

▢ The Meta-Feedback Loop

Create exercises for students to assess AI feedback. Students reflect on the process and share the questions they asked the AI.

Example: Student Critical Thinking Prompt: " You [the AI] corrected my sentence from, 'I have seen him yesterday' to 'I saw him yesterday.'"

- [“Tell the AI that you disagree with the correction and ask for an explanation.”]
  - “I disagree with your correction. Please give me an explanation of your correction.”
- ["Ask the AI what rule it is applying."]
  - “What grammar rule did you apply to correct me?”
- [“Ask the AI how it would explain this to a classmate in Vietnamese.”]
  - “How can you explain your correction to my Vietnamese classmate?”

▢ Context-Rich Prompting

Teachers, don't just ask for grammar exercises. Provide FULL context.

Weak Prompt: "Create a grammar exercise on the present perfect tense."

Strong Prompt: "Create a grammar exercise on the present perfect tense for Vietnamese university students (B1 level) who often confuse it with simple past. Include culturally relevant examples."

Critically Evaluate: "Show me which specific L1 interference patterns you addressed in your exercise. Explain why you chose these examples. Are there any that might confuse Vietnamese learners?"

▢ Demand Self-Assessment

Always ask AI to evaluate its feedback critically.

Example: After the AI provides feedback on student writing, ask

- "Which errors did you prioritize and why?"
- “What errors did you choose to ignore?”
- "How might a Vietnamese [A2, B1, etc.] learner interpret your corrections?"
- “Rate your feedback's clarity from 1 to 10 and justify.”
- “What cultural assumptions are embedded in your corrections?”

More Examples: Simple Prompts You Can Enhance

▢ Pronunciation Feedback Example

"A Vietnamese student pronounces 'think' as 'tink'."

- “Explain the difference and why it's challenging for Vietnamese speakers.”
- “Provide practice progression, step by step.”
- “Critique your approach/method/strategy.”

Your Four P’s:

▢ Writing Feedback Example

"Review this paragraph by a B2 Vietnamese student."

- “Identify and prioritize errors.”
- “Explain your choices based on pedagogical frameworks.”
- Offer examples at different levels of an introduction paragraph. Label the parts.”

Your Four P’s:

▢ Vocabulary Practice Example

"Design a multi-skill vocabulary exercise on environmental protection."

- “Create an A2-B1 level vocabulary list and consider visual ways to address this vocabulary.”
- “Start a list of questions for students to ask using the vocabulary, then prompt them to come up with their own questions.”
- “Adapt the exercise for different reading levels.”

Your Four P’s:

Evaluating AI Feedback Quality

Criterion	Questions to Ask AI	Red Flags to Watch For
Accuracy	"Cite the specific grammar rule you're applying. Is this prescriptive or descriptive? Show me three authoritative sources."	Vague explanations, outdated rules, ignoring language variation
Cultural Sensitivity	"How does this feedback account for Vietnamese communication styles? What assumptions about directness are you making?"	Western-centric examples, ignoring face-saving needs, inappropriate topics
Pedagogical Value	"Is this feedback at the right level? How does it scaffold learning? What's your theory of error correction here?"	Overwhelming corrections, no prioritization, unclear explanations
Adaptability	"How would you modify this for a shy student? A confident one? A student who fears making mistakes?"	One-size-fits-all responses, rigid feedback patterns

The Ultimate Goal

Transform yourself and your students from **passive AI consumers** to **critical AI collaborators** who understand that AI is a tool to be questioned, directed, and improved—not an infallible authority.

Your Next Step: Challenge AI with critical questions and watch how its responses improve. Empower yourself and your students to take control of AI for personal and educational success.

ASSESS (Acronym): Human Skills to Develop

- **Ask AI**: Learn to formulate clear, specific requests
- **Seek Verification**: Build research habits that extend beyond AI interactions
- **Synthesize**: Practice integrating information from multiple sources
- **Evaluate**: Strengthen judgment and decision-making abilities
- **Scrutinize**: Develop critical analysis skills applicable to all information sources
- **Self-reflect**: Develop metacognitive awareness essential for autonomous learning.

The ASSESS framework is a set of human skills we need to develop to be able to interact with other people and technology in critical ways. We can use this framework for lifelong learning and to collaborate effectively with others, including with AI tools. This is framework of active engagement and critical thinking should be practiced continuously.

Source

Claude AI chat 22/07/2025 “What is the ASSIST acronym in AI feedback loops?” [Claude initially responded incorrectly because I asked for the wrong acronym.] “Sorry, is the acronym actually ASSESS?”

Critical Thinking

To assess means to question or to test. As users of high technology, we can use the ASSESS framework responsibly, enhancing our personal learning and our teaching. This acronym reminds us to be active, not passive, users of AI.

- **When was the first time you realized that AI had limits?**
- **Why is it essential to question AI output?**
- **Why should we redirect the system when we have doubts about the AI’s responses?**

Chatbot

"Chatbots are computer programs that simulate human conversation, written or spoken. These days, chatbots are starting to integrate conversational AI, such as natural language processing (NLP), to understand questions even if they aren't grammatically correct and then respond based on data they have collected.

A chatbot may prompt you to ask a question or describe a problem, to which it will either clarify what you said or provide a response. Some chatbots are simple, responding only to the question asked. Some are sophisticated, learning information about you based on data collected and evolving to better assist you over time.

Chatbot type is determined by its capabilities. Declarative chatbots are more basic than predictive chatbots. ***Declarative chatbots*** *perform one function. These chatbots use NLP, defined rules, and machine learning (ML) to generate automated responses when you ask a question. Declarative, or task-oriented chatbots, are most common in customer support and service, and are best when answering commonly asked questions.* ***Predictive chatbots*** *are more sophisticated and personalized than declarative chatbots. Often considered conversational chatbots, or virtual agents, these AI- and data-driven chatbots are much more interactive and aware."*

Source

https://www.coursera.org/articles/what-is-a-chatbot

Critical Thinking

The term "chatbot" has become common, but not necessarily understood. Many people associate the term with ChatGPT without realizing the broad functionality and widespread use of the word. Yes, OpenAI (the creators of ChatGPT) have a chatbot, but theirs isn't the only one.

- **How familiar are you with OpenAI's ChatGPT?**
- **What do you think of when you hear the word "chatbot"?**
- **Which type of chatbot - Declarative or Predictive - have you used the most?**

Feedback Loop

The following definition of Feedback Loop comes from a sociology experiment. The example is an extremely practical one illustrating ***human*** *behavior! I am using this definition on purpose to help explain how* ***machine learning*** *and* ***artificial intelligence*** *need feedback loops to be better at assisting people:*

“[A] feedback loop, [is] a profoundly effective tool for changing behavior. The basic premise is simple. Provide people with information about their actions in real time (or something close to it), then give them an opportunity to change those actions, pushing them toward better behaviors. Action, information, reaction. [...] But the simplicity of feedback loops is deceptive. They are in fact powerful tools that can help people change bad behavior patterns, even those that seem [unmanageable]. Just as important, they can be used to encourage good habits, turning progress itself into a reward. In other words, feedback loops change human behavior.

A feedback loop involves ***four distinct stages***. First comes the ***data***: A behavior must be measured, captured, and stored. This is the evidence stage. Second, the information must be relayed to the individual, [...] in a context that makes it [completely understandable]. This is the ***relevance*** stage. But even compelling information is useless if we don't know what to make of it, so we need a third stage: ***consequence***. The information must illuminate one or more paths ahead. And finally, the fourth stage: ***action***. There must be a clear moment when the individual can recalibrate [‘change’] a behavior, make a choice, and act. Then that action is measured, and the feedback loop can run once more, every action stimulating new behaviors that inch us closer to our goals.”

Source

https://www.wired.com/2011/06/ff-feedbackloop/

Critical Thinking

The four stages of feedback loops used with humans (data, relevance, consequence, action) mirrors feedback loops used with machine learning and artificial intelligence. **The Four P’s: Prompt, Probe, Push back, Polish** can be used to change AI “behavior” so that we see better results over time. See the Critical AI Feedback Loops Guide for more on The Four P’s.

- **Can you think of a time you had to stop doing one thing and had to try something new instead to get a positive outcome?**
- **How did you stop one behavior and start a new one?**
- **What was the outcome of the behavioral change?**

Generative Pre-trained Transformer (GPT)

- **Generative**: Capable of producing content like text and imagery.
- **Pre-trained**: Models already trained on large data sets to accomplish tasks.
- **Transformer**: A deep learning architecture that transforms input into output.

GPT models are used for a variety of tasks, including answering questions, translating languages, and generating text.

Example Sentences

1. GPT is a natural language system that can answer questions and generate text.
2. GPTs are not limited to natural language processing; they can be trained for various tasks.
3. One ethical concern is the potential for GPT models to generate offensive content.

Source

https://www.coursera.org/articles/what-is-gpt?sockid=105aac6eb81062832e73ba45b99b63e5

Critical Thinking

GPTs use AI to have text-based conversations with humans. They respond with text meant to mimic (copy) human speech.

- **How do you decide what is a "good" question?**
- **How did you learn to ask "good" questions?**
- **What do you do in a conversation when you ask a question and the response doesn't seem to match what you asked?**

Large Language Model (LLM)

"Large language models (LLMs) are a type of artificial intelligence (AI) that uses machine learning algorithms to replicate human language. It uses massive data sets to develop its ability to translate languages, predict text, and generate content. As opposed to natural language processing models (NLPs), LLMs train on much larger data sets, allowing it to use a greater number of parameters to become more complex and closer to human language."

Source

https://www.coursera.org/articles/large-language-models

Critical Thinking

Large Language Models (LLMs) are advanced math. The algorithms "learn" human language the more humans use them.

- **What problems can arise when people communicate with other people?**
- **How might those same issues impact a machine attempting to communicate like a human?**

Machine Learning

“Machine learning is the process of using computers to detect patterns in massive datasets and then make predictions based on what the computer learns from those patterns. This makes machine learning **a specific and narrow type of artificial intelligence**. Full artificial intelligence involves machines that can perform abilities we associate with the minds of human beings and intelligent animals, such as perceiving, learning, and problem solving.

All machine learning is based on algorithms. In general, algorithms are sets of specific instructions that a computer uses to solve problems. In machine learning, algorithms are rules for how to analyze data using statistics. Machine learning systems use these rules to identify relationships between data inputs and desired outputs–usually predictions. To get started, scientists give machine learning systems a set of training data. The systems apply their algorithms to this data to train themselves how to analyze similar inputs they receive in the future.”

Source

https://www.energy.gov/science/doe-explainsmachine-learning#:~:text=Machine%20learning%20is%20the%20process,to%20use%20as%20training%20data.

Critical Thinking

Definition: A subfield of AI using algorithms to create self-learning models predicting outcomes and classifying information without human intervention. **Applications**: Suggesting products, predicting stock market fluctuations, and translating text.

- **How ways can you think of in which machines do our “thinking” for us?**
- **What kinds of data processing do you do for yourself on a daily basis?**

Reinforcement Learning from Human Feedback (RLHF)

“Reinforcement Learning from Human Feedback (RLHF) is a popular technique used to align AI systems with human preferences by training them using feedback from people, rather than relying solely on predefined reward functions. Instead of coding every desirable behavior manually (which is often infeasible in complex tasks) RLHF allows models, especially large language models (LLMs), to learn from examples of what humans consider good or bad outputs. This approach is particularly important for tasks where success is subjective or hard to quantify, such as generating helpful and safe text responses. RLHF has become a

cornerstone in building more aligned and controllable AI systems, making it essential for developing AI that behaves in ways humans intend.”

Source

https://blog.ml.cmu.edu/2025/06/01/rlhf-101-a-technical-tutorial-on-reinforcement-learning-from-human-feedback/

Critical Thinking

RLHF enhances the relevance and accuracy of LLMs, especially in chatbots like Google’s Bard and ChatGPT. It uses human feedback for more accurate and efficient human interaction. RLHF is crucial for training generative AI, continuously improving models through human feedback.

- **Do you ever feel rewarded for providing useful feedback (e.g. at work or at school)? How?**

Frequently Asked Questions

Time Management

This method is too time consuming.

- Start small. Apply this method to your personal use of AI to get used to The Four P's and prompt engineering. Use this method on individual exercises in a textbook rather than committing to a full lesson.
- Consider what you want your students to learn and to best address the gap in their knowledge. What are you doing now to address critical thinking and analytical writing? How much time does that take? Why would you expect this process to take longer or require more effort?
- Effective time management usually begins with assessing needs and planning to address them. If using AI more effectively and critically is in your or your students' future, then make time to develop a new set of skills.

How can we avoid getting stuck in the feedback loop, never going anywhere?

- The feedback loop is a circle, yes, but a closed one: although the process is used for critical thinking, a satisfactory response is still a goal. After applying The Four P's, you can stop asking more verifcation questions when your criteria has been addressed. You only stay in the loop when *you* add more criteria.

Prompt Engineering

Can't I just ask the perfect prompt once or modify it to get the perfect answer?

- The point of a feedback loop isn't to get a perfect outcome on the first try. The point of a feedback loop is to apply critical thinking and critical analysis to English language practice to become better at using the language over time. That happens both at the prompt engineering stage *and* during the feedback loop questioning and assessing processes. In short, yes, you could try to write the "perfect" prompt, but you would be denying yourself language control and formation.

AI may still get the data wrong!

- True. However, the process of questioning the AI and requiring the AI to examine its choices helps to eliminate obviously wrong responses. The feedback loop process makes the AI tool verify what it tells you and forces it to look for alternatives it may not have provided in its first response.
- One way to help reduce or eliminate wrong responses is to complete the feedback loop with more than one AI tool. For instance, you can compare the responses to the *same questions* with ChatGPT and Claude, or with MagicSchool and Gemini. You will learn more this way and realize the scope or limits of each tool when you compare two or more sets of responses.

Teachers' Role; Students' Role

Is this method more for teachers than for students?

- No. The critical AI feedback loop is for any language learner who wants to improve their vocabulary, their grammar, and informal or common use of a language.

What is the teacher's role when students already know how to use AI?

- First, we can not assume that frequency of use is the same as critical use and ability to control an AI tool to make more meaningful language. Second, teachers are lifelong students and should always be willing to learn new skills in service of their profession. Finally, teachers have a responsibility to remain resouceful and competent. That doesn't mean that teachers will have to have a high level of knowledge in every new technology that arises! Teachers can develop basic understanding of technologies that are used both socially and academically to offer relevant and current examples to their students.

Students already rely too much on AI, isn't this going to make it worse?

- We don't know how much or how well our students use AI. What we *do* know is that we have methods for training them to use it more meaningfully. Educators can use the same tools that students regularly use, but for different - academic - outcomes.