

## Module 2: Critical thinking

Time required: 45-60 minutes

This module teaches students to treat AI outputs not as facts, but as unverified claims from an unreliable witness that require forensic investigation.

### Learning objectives – help students to:

- Be skeptical of authoritative-sounding AI output
- Apply a systematic protocol to verify sources, dates and logic
- Identify specific types of AI errors (hallucinations, bias, outdated info)
- Understand the reputational risk of using unverified AI content

### Materials:

- Student devices with access to a GenAI tool
- Copies of the Evidence Protocol worksheet (pg. 4)

## Lesson Overview

# THE HUMAN VALIDATOR

## Find the facts and guard against fakes and misinformation

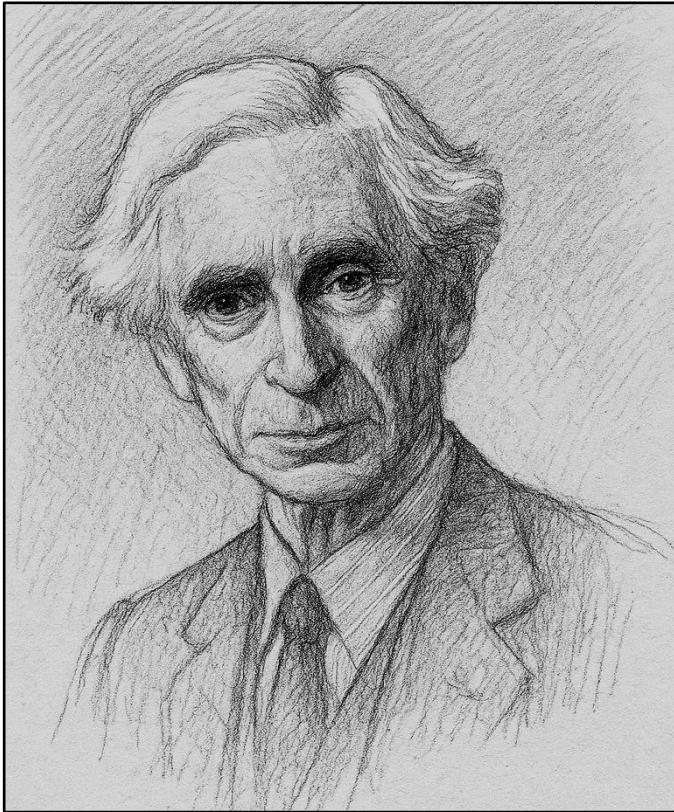


Generative AI is designed to predict plausible text, not to verify facts. It constructs fluent answers based on patterns in its training data, which means it can sound completely confident even when it is completely wrong. Trusting AI without verification is like trusting a rumor because it was told well. Lurking within its output can be misinformation, deepfakes and complete hallucinations.

Helping students develop the skills of the "Human Validator" means urging them to act as forensic investigators. Before building on an AI answer, students must learn to run a credibility check. This is not just about catching errors; it is about taking personal ownership and accountability for the facts.

By verifying sources, triangulating claims and checking for manipulation or bad assumptions, students protect their work, their reputation and the people who will take actions based on their conclusions.

## Bertrand Russell 1872-1970



British logician, mathematician and Nobel laureate, known for analytic philosophy

*“It is undesirable to believe a proposition when there is no ground whatever for supposing it true.”*

From: [“Skeptical Essays”](#) Chapter I: On the value of skepticism  
Published: 1928

- As a leading empiricist, Bertrand Russell argued that knowledge must be grounded in evidence. He rejected beliefs based solely on tradition, intuition or authority. For Russell, a claim is only as strong as the observable facts that support it.
- Russell’s skepticism is the perfect defense against Generative AI models that operate on probability, not experience, generating outputs that can sound logical but lack an empirical basis in the real world.

### Engage the class

**Start the conversation:** “When an AI gives you an answer, do you assume it's true until proven false? Or do you assume it's false until proven true? Russell would say the latter is the only safe setting.”

**Let’s go:** “Today, we are going to stop reading AI outputs as answers and start treating them as ‘evidence from an unreliable witness.’ We are going to learn the forensic skills to spot the lies, the gaps and the hallucinations.”

## EXERCISE

# THE EVIDENCE PROTOCOL

**Introduction** (5 minutes)

Explain the metaphor: You are a detective. The AI is a witness who is confident but prone to making things up. You cannot take their statement to court without verifying it first.

**The investigation** (20 minutes)

Have students prompt an AI to write a short report on a niche or specific historical topic (For example, “Tell me about the most lopsided races for governor in North Carolina history”). Then, have them apply the protocol:

**Track the Source:** Click the links. Do they exist? Do they actually support the claim?

**Check the Dates:** Is the AI relying on older data to answer a question about today?

**Corroborate:** Find a second, non-AI source that confirms the key points.

**Scan for Tampering:** Are there suspicious gaps? Is the tone manipulating you?

**The “busted” analysis** (5 minutes)

Ask students to share what they found. Did anyone find an error? A useless link? A subtle bias?

**Reminder**

Accuracy is not the AI's job; it's yours. If you submit it, you own the error.

## THE EVIDENCE PROTOCOL: Investigator's log

Topic/project: \_\_\_\_\_

### 1. Determine the source (Chain of Custody)

Click the links provided or search for the source text. Does the source actually exist? Notes: \_\_\_\_\_

2. **Check the dates** (Timeliness) When was this evidence created? Is it outdated? Notes: \_\_\_\_\_

### 3. Cross-reference (Corroboration)

Find one external, reputable source (book, academic journal, news site) that agrees or disagrees.

Source found: \_\_\_\_\_ Verdict: [  ] Matches [  ] Contradicts

### 4. Scan for tampering (Omissions/Bias)

What is missing? Whose perspective is left out? Is the language biased? Notes: \_\_\_\_\_

### 5. Recompute the details (Precision)

Pick one number, quote or specific fact. Verify it exactly. Notes: \_\_\_\_\_

### FINAL VERDICT:

Based on the evidence, is this AI output admissible?

[  ] Admissible (Verified) [  ] Inadmissible (Flawed/False)

What specific correction must be made? \_\_\_\_\_

1

Bertrand Russell argues it is "undesirable" to believe things without grounds. Why is this harder to do with AI than with a Google search? Does the conversational, confident tone of a chatbot make us lower our guard?

2

In the exercise, did anyone find a "hallucination" (a made-up fact)? How plausible did it look before you checked it? What would have happened if you used that fact in a paper or a job presentation?

3

The guide suggests checking for "what's not being said." Why is AI (which is trained on certain categories of online content) prone to leaving out minority viewpoints or controversial counter-arguments?

4

If you are the "Human Validator," does that change how you view your value? Instead of being a "creator" of content, are you becoming an "editor" or "judge" of truth? Is that a promotion or a demotion?

5

If the internet becomes populated with AI-generated output that may contain errors, what are the implications for the future of truth and facts and what are the implications for society?

### Reflective assessment

How well did this module enable students to:

- Treat AI output with skepticism?
- Identify specific hallucinations or errors in a text?
- Understand the reputational stakes of using AI without fact-checking
- Think about the future of truth?

How can you build on this module to help students develop a habit of verification?



# HUMAN WISDOM FOR THE AGE OF AI

A FIELD GUIDE TO CULTIVATING ESSENTIAL SKILLS

AI-U 2026

[www.studentguidetoai.org](http://www.studentguidetoai.org)

## Imagining the Digital Future Center

Elon University

Elon, North Carolina, USA

[imagine@elon.edu](mailto:imagine@elon.edu)

[imaginingthedigitalfuture.org](http://imaginingthedigitalfuture.org)

© 2026 Elon University. This work is licensed under Creative Commons [CC-BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/). You are free to share and adapt this work for noncommercial purposes, but attribution to Elon University is required.



Imagining the  
Digital Future  
Center

