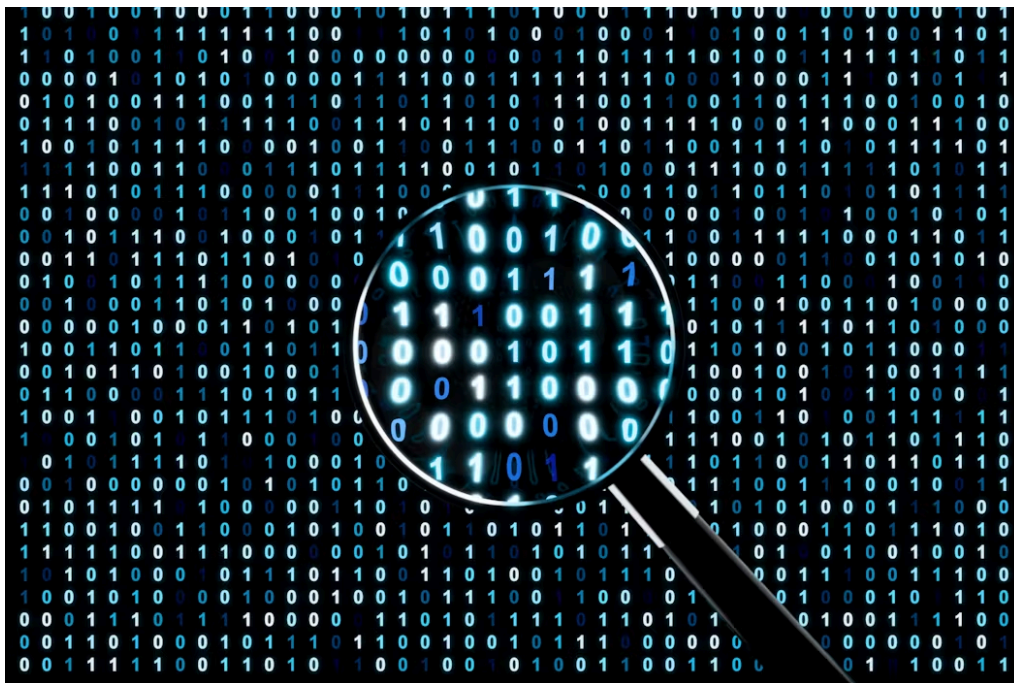


# New technologies like AI come with big claims – borrowing the scientific concept of validity can help cut through the hype

Kai R. Larsen, Professor of Information Systems, University of Colorado Boulder

Published: November 12, 2025 8:36am EDT



Closely examining the claims companies make about a product can help you separate hype from reality.

*Flavio Coelbo/Moment via Getty Images*

Technological innovations can seem relentless. In computing, some have proclaimed that “a year in machine learning is a century in any other field.” But how do you know whether those advancements are hype or reality?

Failures quickly multiply when there's a deluge of new technology, especially when these developments haven't been properly tested or fully understood. Even technological innovations from trusted labs and organizations sometimes result in spectacular failures. Think of IBM Watson, an AI program the company hailed as a revolutionary tool for cancer treatment in 2011. However, rather than evaluating the tool based on patient outcomes, IBM used less relevant measures – possibly even irrelevant ones, such as expert ratings rather than patient outcomes. As a result, IBM Watson not only failed to offer doctors reliable and innovative treatment recommendations, it also suggested harmful ones.

When ChatGPT was released in November 2022, interest in AI expanded rapidly across industry and in science alongside ballooning claims of its efficacy. But as the vast majority of companies are seeing their attempts at incorporating generative AI fail, questions about whether the technology does what developers promised are coming to the fore.



IBM Watson wowed on Jeopardy, but not in the clinic.

*AP Photo/Seth Wenig*

In a world of rapid technological change, a pressing question arises: How can people determine whether a new technological marvel genuinely works and is safe to use?

Borrowing from the language of science, this question is really about validity – that is, the soundness, trustworthiness and dependability of a claim. Validity is the ultimate verdict of whether a scientific claim accurately reflects reality. Think of it as quality control for science: It helps researchers know whether a medication really cures a disease, a health-tracking app truly improves fitness, or a model of a black hole genuinely describes how it behaves in space.

How to evaluate validity for new technologies and innovations has been unclear, in part because science has mostly focused on validating claims about the natural world.

In our work as researchers who study how to evaluate science across disciplines, we developed a framework to assess the validity of any design, be it a new technology or policy. We believe setting clear and consistent standards for validity and learning how to assess it can empower people to make informed decisions about technology – and determine whether a new technology will truly deliver on its promise.

## **Validity is the bedrock of knowledge**

Historically, validity was primarily concerned with ensuring the precision of scientific measurements, such as whether a thermometer correctly measures temperature or a psychological test accurately assesses anxiety. Over time, it became clear that there is more than just one kind of validity.

Different scientific fields have their own ways of evaluating validity. Engineers test new designs against safety and performance standards. Medical researchers use controlled experiments to verify treatments are more effective than existing options.

Researchers across fields use different types of validity, depending on the kind of claim they're making.

Internal validity asks whether the relationship between two variables is truly causal. A medical researcher, for instance, might run a randomized controlled trial to be sure that a new drug led patients to recover rather than some other factor such as the placebo effect.

External validity is about generalization – whether those results would still hold outside the lab or in a broader or different population. An example of low external validity is how many early studies that work in mice don't always translate to people.

Construct validity, on the other hand, is about meaning. Psychologists and social scientists rely on it when they ask whether a test or survey really captures the idea it's supposed to measure. Does a grit scale actually reflect perseverance or just stubbornness?

Finally, ecological validity asks whether something works in the real world rather than just under ideal lab conditions. A behavioral model or AI system might perform brilliantly in simulation but fail once human behavior, noisy data or institutional complexity enter the picture.

Across all these types of validity, the goal is the same: ensuring that scientific tools – from lab experiments to algorithms – connect faithfully to the reality they aim to explain.

# Evaluating technology claims

We developed a method to help researchers across disciplines clearly test the reliability and effectiveness of their inventions and theories. The design science validity framework identifies three critical kinds of claims researchers usually make about the utility of a technology, innovation, theory, model or method.

First, a criterion claim asserts that a discovery delivers beneficial outcomes, typically by outperforming current standards. These claims justify the technology's utility by showing clear advantages over existing alternatives.

For example, developers of generative AI models such as ChatGPT may see higher engagement with the technology the more it flatters and agrees with the user. As a result, they may program the technology to be more affirming – a feature called sycophancy – in order to increase user retention. The AI models meet the criterion claim of users considering them more flattering than talking to people. However, this does little to improve the technology's efficacy in tasks such as helping resolve mental health issues or relationship problems.

Second, a causal claim addresses how specific components or features of a technology directly contribute to its success or failure. In other words, it is a claim that shows researchers know what makes a technology effective and exactly why it works.

Looking at AI models and excessive flattery, researchers found that interacting with more sycophantic models reduced users' willingness to repair interpersonal conflict and increased their conviction of being in the right. The causal claim here is that the AI feature of sycophancy reduces a user's desire to repair conflict.

Third, a context claim specifies where and under what conditions a technology is expected to function effectively. These claims explore whether the benefits of a technology or system generalize beyond the lab and can reach other populations and settings.

In the same study, researchers examined how excessive flattery affected user actions in other datasets, including the "Am I the Asshole" community on Reddit. They found that AI models were more affirming of user decisions than people were, even when the user was describing manipulative or harmful behavior. This supports the context claim that sycophantic behavior from an AI model applies across different conversational contexts and populations.

# Measuring validity as a consumer

Understanding the validity of scientific innovations and consumer technologies is critical for scientists and the general public. For scientists, it's a road map to ensure their inventions are rigorously evaluated. And for the public, it means knowing that the tools and systems they depend on – such as health apps, medications and financial platforms – are truly safe, effective and beneficial.

Here's how you can use validity to understand the scientific and technological innovations happening around you.

Because it is difficult to compare every feature of two technologies against each other, focus on which features you value most from a technology or model. For example, do you prefer a chatbot to be accurate or better for privacy? Examine claims for it in that area, and check that it is as good as claimed.

Consider not only the types of claims made for a technology but also which claims are not made. For example, does a chatbot company address bias in its model? It's your key to knowing whether you see untested and potentially unsafe hype or a genuine advancement.

By understanding validity, organizations and consumers can cut through the hype and get to the truth behind the latest technologies.

The authors do not work for, consult, own shares in or receive funding from any company or organization that would benefit from this article, and have disclosed no relevant affiliations beyond their academic appointment.

This article is republished from *The Conversation* under a Creative Commons license.