

A checklist of eighteen pitfalls in AI journalism

Sayash Kapoor, Arvind Narayanan

September 30, 2022

Read the blog post which introduces this checklist [here](#).

Flawed human-AI comparison

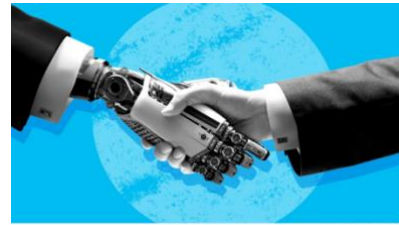
What? A false comparison between AI tools and humans that implies AI tools and humans are similar in how they learn and perform.

Why is this an issue? Rather than describing AI as a broad set of tools, such comparisons anthropomorphize AI tools and imply that AI tools have the potential to act as agents in the real world.

Pitfall 1. Attributing agency to AI: Describing AI systems as taking actions independent of human supervision or implying that they may soon be able to do so.

“Artificial intelligence is starting to take over repetitive tasks in classrooms, like grading”
—[The Machine Are Learning, and So Are the Students](#), The New York Times

Pitfall 2. Suggestive imagery: Images of humanoid robots are often used to illustrate articles about AI, even if the contents of the article have nothing to do with robots. This gives readers a false impression that AI tools are embodied, even when it is just software that learns patterns from data.



—[How elite investors use artificial intelligence and machine learning to gain an edge](#), CNN

This article has nothing to do with robots and is about AI tools to find patterns in financial data.

Pitfall 3. Comparison with human intelligence: In some cases, articles on AI imply that AI algorithms learn in the same way as humans do. For example, comparisons of deep learning algorithms with the way the human brain functions are common. Such comparisons can lend credence to claims that AI is “sentient”, as Dr. Timnit Gebru and Dr. Margaret Mitchell note in their [recent op-ed](#).

“[The study] focused on an AI technique called deep learning, which employs algorithms, big data, and computing power to emulate human intelligence.”

—[AI may be as effective as medical specialists at diagnosing disease](#), CNN

“Emulating human intelligence” is not an accurate description of what deep learning does. It gives readers a false intuition that deep learning algorithms compare with human intelligence.

Pitfall 4. Comparison with human skills: Similarly, articles often compare how well AI tools perform with human skills on a given task. This falsely implies that AI tools and humans compete on an equal footing—hiding the fact that AI tools only work in a narrow range of settings.

“A new scientific review has concluded that artificial intelligence (AI) may be able to diagnose disease as successfully as human healthcare professionals”

—[AI may be as effective as medical specialists at diagnosing disease](#), CNN

This sentence hides the fact that AI tools only perform a narrow slice of the variety of steps that comprise a diagnosis.

Hyperbolic, incorrect, or non-falsifiable claims about AI

What? Claims about AI tools that are speculative, sensational, or incorrect can spread hype about AI.

Why is this an issue? Such claims give a false sense of progress in AI and make it difficult to identify where true advances are being made.

Pitfall 5. Hyperbole: Describing AI systems as revolutionary or groundbreaking without concrete evidence of their performance gives a false impression of how useful they will be in a given setting. This issue is amplified when the AI tool is deployed in a setting where they are known to have past failures—we should be skeptical about the effectiveness of AI tools in these settings.

*“For years, people have tried to re-engineer learning with artificial intelligence, **but it was not until the machine-learning revolution** of the past seven years that real progress has been made.” (emphasis ours)*

—[The Machine Are Learning, and So Are the Students](#), The New York Times

This statement has no evidence in the article to back it up. In fact, EdTech has proven to be notoriously [failure-prone](#) in the last decade.

Pitfall 6. Uncritical comparison with historical transformations: Comparing AI tools with major historical transformations like the invention of

“In Altman’s view, the unfolding AI revolution may well be more consequential for humanity than the

electricity or the industrial revolution is a great marketing tactic. However, when news articles adopt these terms, they can convey a false sense of potential and progress—especially when these claims are not backed by real-world evidence.

preceding agricultural, industrial and computer revolutions combined.”

—[Is AI finally closing in on human intelligence?](#), Financial Times

The article uncritically quotes Sam Altman, co-founder of OpenAI, in comparing AI with historical transformations.

Pitfall 7. Unjustified claims about future

progress: Claims about how future developments in AI tools will affect an industry, for instance, by implying that AI tools will inevitably be useful in the industry. When these claims are made without evidence, they are mere speculation on the part of the article, and like before, can give a false impression about these developments.

“Chatbots, for example, can be clumsy and frustrating today, but they will eventually become truly conversational, learning our habits and personalities and even develop personalities of their own.”

—[A.I. Here, There, Everywhere](#), The New York Times

Pitfall 8. False claims about progress: In some cases, articles can include false claims about what an AI tool can do.

“The system’s auto-grader teaches itself how to score.”

—[The Machine Are Learning, and So Are the Students](#), The New York Times

The article falsely claims that the tools can learn how to grade homework on their own, when in fact the tool merely scores student answers against the correct answers fed into the system.

Pitfall 9. Incorrect claims about what a study reports: News articles often cite academic studies to substantiate their claims. Unfortunately, there is sometimes a gap between the claims made based on an academic study and what the study reports.

“[Studies show](#) that [ML] systems can raise student performance well beyond the level of conventional classes and even beyond the level achieved by students who receive instruction from human tutors.”

—[The Machine Are Learning, and So Are the Students](#), The New York Times

The study cited here does not refer to machine learning even once.

Pitfall 10. Deep-sounding terms for banal

actions: As Prof. Emily Bender discusses in her work on dissecting AI hype, using phrases like “[the elemental act of next-word prediction](#)” or “[the magic of AI](#)” implies that an AI tool is doing something remarkable in the course of its operation. It hides an understanding of how mundane the tasks are, and that AI tools are functioning exactly as expected.

“I ask the gods of artificial intelligence to turn on the light”

—[A.I. Here, There, Everywhere](#), The New York Times

Uncritically platforming those with self-interest

What? News articles often use PR statements and quotes from company spokespeople to substantiate their claims without providing adequate context or balance in their news stories.

Why is this an issue? Emphasizing the opinions of self-interested parties without providing alternative viewpoints can give an over-optimistic sense of progress.

Pitfall 11. Treating company spokespeople and

researchers as neutral parties: When an article only or primarily has quotes from company spokespeople or researchers who built an AI tool, it is likely to be over-optimistic about the potential benefits of the tool.

[AI tested as university exams undergo digital shift](#), Financial Times

Almost the entire article is written from the perspective of the company selling AI tools. As a result, the article reads more like a PR piece and less like a news story.

Pitfall 12. Repeating or re-using PR terms and

statements: News articles often re-use terms from companies’ PR statements instead of describing how an AI tool works. This can lead to misleading wording that misrepresents the actual capabilities of a tool.

“She uses the platform Bakpax that can read students’ handwriting and auto-grade schoolwork”

—[The Machine Are Learning, and So Are the Students](#), The New York Times

The article repeatedly reuses PR terms such as “read students’ handwriting” and “auto-grade” homework. Though Bakpax has since shut down, we found these PR terms on the company’s [archived website](#).

Limitations not addressed

What? The potential benefits of an AI tool are emphasized, but the potential limitations are not addressed or emphasized.

Why is this an issue? A one-sided analysis of AI tools can hide the potential limitations of these tools.

Pitfall 13. No discussion of potential limitations:

Limitations such as inadequate validation, bias, and potential for dual-use plague most AI tools. When these limitations are not discussed, readers can get a skewed view of the risks associated with AI tools.

[The Machine Are Learning, and So Are the Students](#), The New York Times

There is no discussion of potential limitations of the use of AI in EdTech. The article mentions in passing that there could be privacy concerns, but quickly provides a quote from the developer of the tool to assuage these concerns.

Pitfall 14. Limitations de-emphasized: Even if an article discusses these limitations and quotes experts who can explain them, limitations are often downplayed in the structure of the article, for instance by positioning them at the end of the article or giving them limited space.

“The big caveat is, in my opinion, that the story is not ‘AI may be as good as health professionals’, but that ‘the general standard of evaluating performance of AI is shoddy.’”

—[AI may be as effective as medical specialists at diagnosing disease](#), CNN

In this story about a research study, despite an expert taking pains to point out what the real message of the study is, the article buries this information at the far end.

Pitfall 15. Limitations addressed in a “skeptics” framing: Limitations of AI tools can be caveated in the framing of the article by positioning experts who explain these limitations as skeptics who don’t see the true potential of AI.

“Some skeptics argue that the software is capable only of blind mimicry ...”

—[A.I. Is Mastering Language. Should We Trust What It Says?](#), The New York Times Magazine

Instead of engaging with the substantive points of Prof. Emily Bender and others, this NYT Mag piece circumscribed their viewpoints to a skeptics’ framing. Prof. Bender discusses issues with this article in [much more detail](#) in her response to this article.

Pitfall 16. Downplaying human labor: When discussing AI tools, articles often foreground the role of technical advances and downplay all the human labor that is necessary to build the system or keep it running. The book [Ghost Work](#) by Dr. Mary L. Gray and Dr. Siddharth Suri reveals how important this invisible labor is. Downplaying human labor misleads readers into thinking that AI tools work autonomously, instead of clarifying that they require [significant overhead](#) in terms of human labor, as Prof. Sarah T. Roberts discusses.

“A plethora of online courses and tutorials also have freed teachers from lecturing”

—[The Machine Are Learning, and So Are the Students](#), The New York Times

This phrase does not acknowledge the labor that goes into recording these lectures, maintaining online resources, and shifts the attention away from the human labor that goes into maintaining this system.

Pitfall 17. Performance numbers reported without uncertainty estimation or caveats: There is seldom enough space in a news article to explain how performance numbers like accuracy are calculated for a given application or what they represent. Including numbers like “90% accuracy” in the body of the article without specifying the conditions under which these numbers are calculated can misinform readers about the efficacy of an AI tool, especially because AI tools are known to suffer performance degradations even under slight changes to the datasets they are evaluated on.

“A new computer algorithm can now forecast crime in a big city near you — apparently. The algorithm, which was formulated by social scientists at the University of Chicago and touts 90% accuracy...”

—[Algorithm Claims to Predict Crime in US Cities Before It Happens](#), Bloomberg

The article has no details that might help the reader to understand what an accuracy of 90% means.

Pitfall 18. The fallacy of inscrutability: Referring to AI tools as inscrutable black boxes [is a category error](#). Instead of holding the developers of these tools accountable for their design choices, it shifts scrutiny to the technical aspects of the system. Journalists should hold developers accountable for the performance of AI tools rather than referring to these tools as black boxes and allowing developers to evade accountability.

[Our Machines Now Have Knowledge We'll Never Understand](#), WIRED

The article's main point is that it is impossible to understand how models “reason”, simply because they have a large number of parameters or weights. It ignores an [entire body of research](#) on model interpretability and explainability. The author further argues that because we cannot understand the model's internal representations, there is no way to use them in a way that meets legal requirements for non-discrimination and explanation, such as in credit scoring. But those requirements are about the way that decision making algorithms interact with the world rather than their internals, and

algorithms can always be understood at this level, as Dr. Kroll [notes](#).

Revisions:

- September 30th: Added citations to related work, changed the example in pitfall 18, corrected the text in pitfall 7.