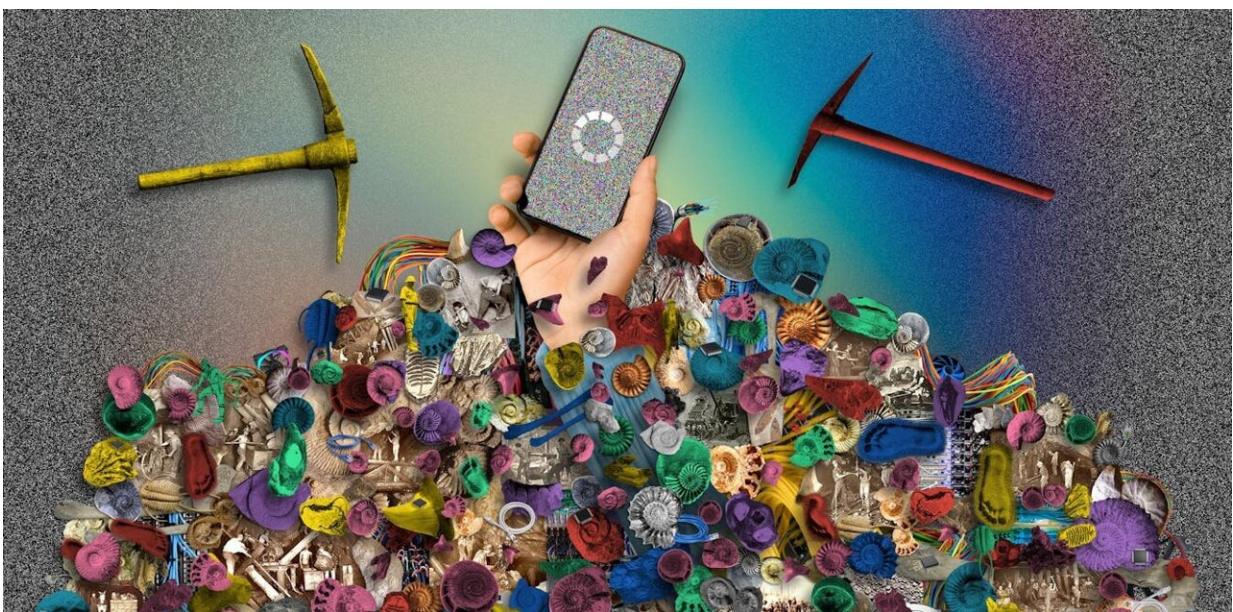


Viewpoint: Generative AI hype is ending—and now the technology might actually become useful

August 19 2024, by Vitomir Kovanovic



Credit: [Luke Conroy and Anne Fehres & AI4Media / Better Images of AI / Models Built From Fossils, CC BY](#)

Less than two years ago, the launch of ChatGPT started a generative AI frenzy. Some said the technology would trigger a [fourth industrial revolution](#), completely reshaping the world as we know it.

In March 2023, Goldman Sachs predicted [300 million jobs would be lost](#)

or degraded due to AI. A huge shift seemed to be underway.

Eighteen months later, [generative AI is not transforming business](#). Many projects using the [technology](#) are being cancelled, such as an attempt by McDonald's to [automate drive-through ordering](#) which went viral on TikTok after [producing comical failures](#). Government efforts to make systems to [summarize public submissions](#) and [calculate welfare entitlements](#) have met the same fate.

So what happened?

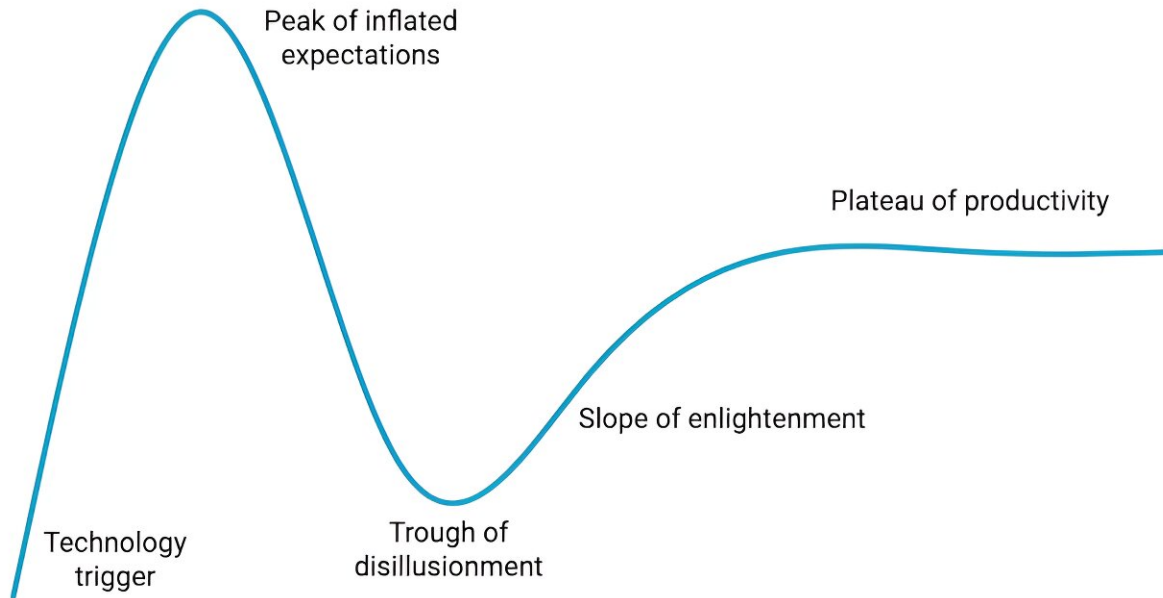
The AI hype cycle

Like many new technologies, generative AI has been following a path known as the Gartner hype cycle, first described by American tech research firm Gartner.

This widely used model describes a recurring process in which the initial success of a technology leads to inflated public expectations that eventually fail to be realized. After the early "peak of inflated expectations" comes a "trough of disillusionment," followed by a "slope of enlightenment" which eventually reaches a "plateau of productivity."

The Gartner hype cycle

A model of how hype around new technologies develops over time.



Credit: The Conversation

A [Gartner report published in June](#) listed most generative AI technologies as either at the peak of inflated expectations or still going upward. The report argued most of these technologies are two to five years away from becoming fully productive.

Many compelling prototypes of generative AI products have been developed, but adopting them in practice has been less successful. A [study published last week](#) by American think tank RAND showed 80% of AI projects fail, more than double the rate for non-AI projects.

Shortcomings of current generative AI technology

The RAND report lists many difficulties with generative AI, ranging from high investment requirements in data and AI infrastructure to a lack of needed human talent. However, the unusual nature of GenAI's limitations represents a critical challenge.

For example, generative AI systems can solve some highly complex university admission tests yet fail [very simple tasks](#). This makes it very hard to judge the potential of these technologies, which leads to false confidence.

After all, if it can solve complex differential equations or write an essay, it should be able to take simple drive-through orders, right?

A [recent study](#) showed that the abilities of large language models such as GPT-4 do not always match what people expect of them. In particular, more capable models severely underperformed in high-stakes cases where incorrect responses could be catastrophic.

These results suggest these models can induce false confidence in their users. Because they fluently answer questions, humans can reach overoptimistic conclusions about their capabilities and deploy the models in situations they are not suited for.

Experience from successful projects shows it is tough to make a generative model follow instructions. For example, Khan Academy's Khanmigo tutoring system often revealed the correct answers to questions despite being instructed not to.

So why isn't the generative AI hype over yet?

There are a few reasons for this.

First, generative AI technology, despite its challenges, is rapidly

improving, with scale and size being the primary drivers of the improvement.

[Research shows](#) that the size of language models (number of parameters), as well as the amount of data and computing power used for training all contribute to improved model performance. In contrast, the architecture of the neural network powering the model seems to have minimal impact.

Large language models also display so-called [emergent abilities](#), which are unexpected abilities in tasks for which they haven't been trained. Researchers have [reported](#) new capabilities "emerging" when models reach a specific critical "breakthrough" size.

Studies have found sufficiently complex large language models can develop the ability to [reason by analogy](#) and even [reproduce optical illusions](#) like those experienced by humans. The precise causes of these observations are [contested](#), but there is no doubt large language models are becoming more sophisticated.

So AI companies are still at work on bigger and more expensive models, and [tech companies](#) such as Microsoft and Apple are betting on returns from their existing investments in generative AI. According to [one recent estimate](#), generative AI will need to produce US\$600 billion in annual revenue to justify current investments—and this figure is likely to [grow to US\\$1 trillion](#) in the coming years.

For the moment, the biggest winner from the generative AI boom is Nvidia, the largest producer of the chips powering the generative AI arms race. As the proverbial shovel-makers in a gold rush, Nvidia recently became the most valuable public company in history, tripling its share price in a single year to reach [a valuation of US\\$3 trillion](#) in June.

What comes next?

As the AI hype begins to deflate and we move through the period of disillusionment, we are also seeing more realistic AI adoption strategies.

First, AI is being used to support humans, rather than replace them. A [recent survey of American companies](#) found they are mainly using AI to improve efficiency (49%), reduce labor costs (47%) and enhance the quality of products (58%)

Second, we also see a [rise in smaller \(and cheaper\) generative AI models](#), trained on specific data and deployed locally to reduce costs and optimize efficiency. Even OpenAI, which has led the race for ever-larger models, has released the GPT-4o Mini [model](#) to reduce costs and improve performance.

Third, we see a strong [focus on providing AI literacy training](#) and educating the workforce on how AI works, its potentials and limitations, and best practices for ethical AI use. We are likely to have to learn (and re-learn) how to use different AI technologies for years to come.

In the end, the AI revolution will look more like an evolution. Its use will gradually grow over time and, little by little, alter and transform human activities. Which is much better than replacing them.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Viewpoint: Generative AI hype is ending—and now the technology might actually become useful (2024, August 19) retrieved 3 September 2024 from

<https://techxplore.com/news/2024-08-viewpoint-generative-ai-hype-technology.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.