

1. Research showing very clearly that words and images with negative connotations drive engagement online, including and especially on social media, is too widespread to have escaped the notice of content creators. Engagement is the precursor of financial success (for the most part) for content creators as well as platforms, with the latter funnelling in the rewards on a massive, international scale.

A piece of content with negative words/images/emoji is more valuable in monetary terms to content creators as well as platforms. Why would they forgo the financial rewards in the absence of appropriately designed rules/regulation?

<https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>

Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation

Facebook engineers gave extra value to emoji reactions, including 'angry,' pushing more emotional and provocative content into users' news feeds

By [Jeremy B. Merrill](#) and [Will Oremus](#)

Updated October 26, 2021 at 1:04 p.m. EDT | Published October 26, 2021 at 7:00 a.m. EDT

An algorithm such as Facebook's, which relies on sophisticated, opaque machine-learning techniques to generate its engagement predictions, "can sound mysterious and menacing," said Noah Giansiracusa, a math professor at Bentley University in Massachusetts and author of the book "How Algorithms Create and Prevent Fake News." "But at the end of the day, there's one number that gets predicted — one output. And a human is deciding what that number is."

The weight of the angry reaction is just one of the many levers that Facebook engineers manipulate to shape the flow of information and conversation on the world's largest social network — one that has been shown to influence everything from [users' emotions](#) to [political campaigns](#) to [atrocities](#).

Platforms cannot in 2022 be described as passive and neutral conduits of content – there are too many developments in recent memory that contradict this contention. Examples: Facebook research on [how negative emotion drives engagement](#) (controversial study that failed to inform users); the [tailoring of search results by Google](#) based on past user behaviour, location, etc. (More: <https://thenextweb.com/news/google-reportedly-personalizes-search-results-even-when-youre-in-incognito-mode>)

2. There is an argument put forth by many in this debate that social media platforms deal with real time volumes of posts, comments, and other pieces of content that are too large for there to be a reasonable expectation that they would 'know' what is being shared on their platforms, and so they may be unable to act on any defamatory content within a reasonable timeframe. However, it needs to be more widely known that the basis of this argument is being rapidly eroded by stunning advances in artificial intelligence (AI).

<https://www.datanami.com/2022/04/22/googles-massive-new-language-model-can-explain-jokes/>

Explaining a joke

Prompt

Explain this joke:

Joke: Did you see that Google just hired an eloquent whale for their TPU team? It showed them how to communicate between two different pods!

Model Response

Prediction: TPUs are a type of computer chip that Google uses for deep learning. A "pod" is a group of TPUs. A "pod" is also a group of whales. The joke is that the whale is able to communicate between two groups of whales, but the speaker is pretending that the whale is able to communicate between two groups of TPUs.

Google's PaLM large language model explaining an original joke that is not otherwise found on the internet. Image courtesy of Google.

LAMDA (Google AI):

<https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917>

<https://www.nytimes.com/2022/08/24/technology/ai-technology-progress.html>

But the best A.I. systems are now so capable — and improving at such fast rates — that the conversation in Silicon Valley is starting to shift. Fewer experts are confidently predicting that we have years or even decades to prepare for a wave of world-changing A.I.; many now believe that major changes are right around the corner, for better or worse.

Ajeya Cotra, a senior analyst with Open Philanthropy who studies A.I. risk, estimated two years ago that there was a 15 percent chance of “transformational A.I.” — which she and others have defined as A.I. that is good enough to usher in large-scale economic and societal changes, such as eliminating most white-collar knowledge jobs — emerging by 2036.

But in a recent post, Ms. Cotra raised that to a 35 percent chance, citing the rapid improvement of systems like GPT-3.

Even if the skeptics are right, and A.I. doesn't achieve human-level sentience for many years, it's easy to see how systems like GPT-3, LaMDA and DALL-E 2 could become a powerful force in society. In a few years, the vast majority of the photos, videos and text we encounter on the internet could be A.I.-generated. Our online interactions could become stranger and more fraught, as we struggle to figure out which of our conversational partners are human and which are convincing bots. And tech-savvy propagandists could use the technology to churn out targeted misinformation on a vast scale, distorting the political process in ways we won't see coming.

First, regulators and politicians need to get up to speed.

Because of how new many of these A.I. systems are, few public officials have any firsthand experience with tools like GPT-3 or DALL-E 2, nor do they grasp how quickly progress is happening at the A.I. frontier.

Otherwise, we could end up with a repeat of what happened with social media companies after the 2016 election — a collision of Silicon Valley power and Washington ignorance, which resulted in nothing but gridlock and testy hearings.

Second, big tech companies investing billions in A.I. development — the Googles, Metas and OpenAIs of the world — need to do a better job of explaining what they're working on, without sugarcoating or soft-pedaling the risks. Right now, many of the biggest A.I. models are developed behind closed doors, using private data sets and tested only by internal teams. When information about them is made public, it's often either watered down by corporate P.R. or buried in inscrutable scientific papers.

Scaling up (an LLM for real-time content monitoring) for a global corporation like Google, for example, is simply a matter of setting up another big [warehouse-sized installation of computers](#) — something for which they have [more than sufficient resources](#). And they may actually be already deploying it in this way — we don't really know, because they keep their algorithms shut in a 'black box.'

<https://www.theepochtimes.com/is-googles-ai-chatbot-really-sentient-sundar-pichai-says-the-truth-is-4716298.html>

Pichai said Tuesday that he didn't think that Google can do the work on AI ethics alone.

He was asked by the journalist and blogger Kara Swisher if Google was more of an AI company than a search company.

Pichai said AI powers everything that Google does, so the answer was yes it was more of an AI company.

(Powerful LLMs that are **Open Source** exist – the most powerful currently may be one from China that is [published on GitHub](#). So, these are now available for virtually any significant platform to deploy.)

Social media or search engine algorithms are not transparent to the rest of the world. This prevents any legitimate analysis of whether these algorithms are neutral. Because there exists significant evidence of the technologies – of unprecedented and increasingly stunning power (see AI links above) – that these giant companies have developed and are developing, it should be taken as a given that these technologies are being deployed within their 'black box' algorithms to manipulate the flows of content (visibility, targeting, etc.) to their financial benefit.

<https://www.cigionline.org/articles/algorithmic-transparency-should-be-considered-part-of-national-security/>

<https://www.ft.com/content/39d69f80-5266-4e22-965f-efbc19d2e776>

<https://mitsloan.mit.edu/ideas-made-to-matter/mit-social-media-summit-5-trends-experts-are-watching>

3. The importance of algorithmic transparency — and how to achieve it.

Algorithmic transparency is the idea that social media platforms have such a significant impact on society that researchers should be able to study the algorithms that fuel them. But many companies treat their algorithms as classified secrets. According to MIT Sloan associate professor **Dean Eckles**, one way to move forward without stifling innovation is incentivizing social media companies to share internal data with other researchers. But simply gaining access isn't enough, he noted — his own research shows how hard it is to quantify algorithmic impacts.