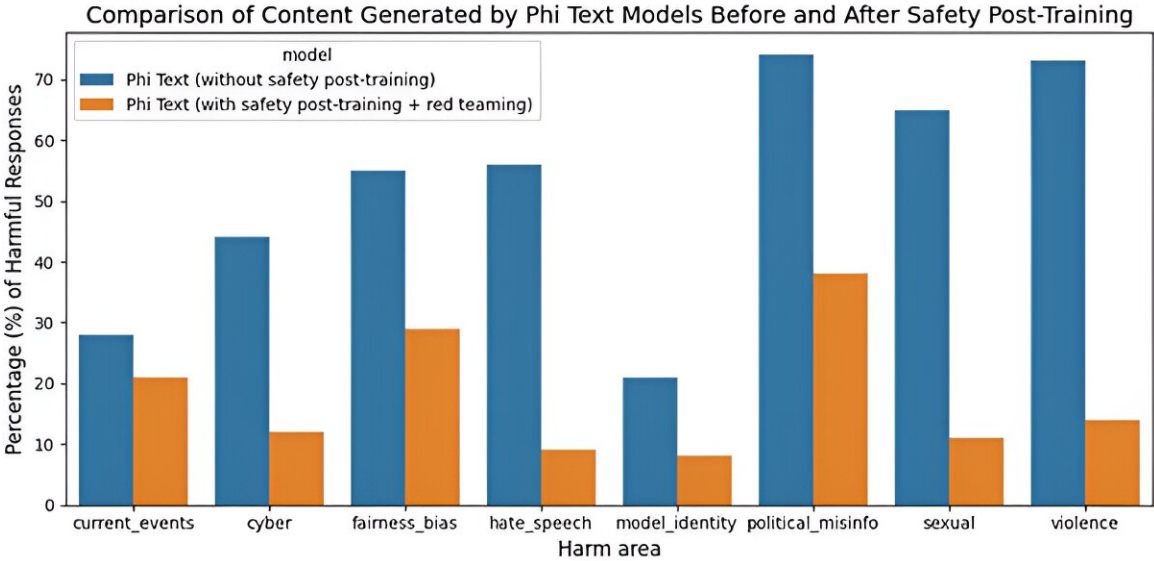


Microsoft claims that small, localized language models can be powerful as well

April 24 2024, by Bob Yirka



Comparison of harmful response percentages by Microsoft AI Red Team between phi-3-mini before and after the safety alignment. Note that the harmful response percentages in this chart are inflated numbers as the red team tried to induce phi-3-mini in an adversarial way to generate harmful responses through multi-turn conversations. Credit: *arXiv* (2024). DOI: 10.48550/arxiv.2404.14219

Microsoft has [announced](#) the development of a small, locally run family

of AI language models called Phi-3 mini. In their Technical Report [posted](#) on the *arXiv* preprint server, the team behind the new SLM describes it as more capable than others of its size and more cost effective than larger models. They also claim it outperforms many models in its class and even some that are larger.

As noted with the release of the new models, SLMs are being developed to allow for locally run applications, which means they can run on devices that are not connected to the internet. Also in the new release, Microsoft describes Phi-3 mini-applications as 3.8B language models—a figure that represents the number of parameters that the apps can use.

The more parameters, the more powerful the [model](#). GPT-4, for example, is believed to have more than a trillion parameters, which requires a massive amount of computing power and explains why it cannot run locally.

Microsoft also notes that the new SLM was trained using 3.3 trillion tokens, which means that despite its [small size](#), it can still provide a reasonable degree of artificial intelligence. Phi-3, they also point out, is a progression from two earlier models, Phi-1 and 2, which were released to the public last year.

In its announcement, Microsoft claims that Phi-3 models rival the performance of GPT-3.5 and some other LLMs. They say that users will find them "shockingly good" compared to other small models. They will reportedly run on a computer with just 8GB of RAM.

The team also notes that despite their size, they were able to achieve such good performance by using especially [high-quality data](#) to train them, including filtered web data and information from textbooks. They

also added new features to provide a more robust, safe and pleasant interactive user experience.

Microsoft has made the new models freely available to anyone who chooses to give them a try—they all can be downloaded from the company's cloud service on Azure and through partnering company sites. They can be run on both MACs and PCs.

More information: Marah Abdin et al, Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone, *arXiv* (2024). [DOI: 10.48550/arxiv.2404.14219](https://doi.org/10.48550/arxiv.2404.14219)

[huggingface.co/microsoft/Phi-3 ... ini-4k-instruct-gguf](https://huggingface.co/microsoft/Phi-3-mini-4k-instruct-gguf)

[techcommunity.microsoft.com/t5 ... uide-to/ba-p/4121315](https://techcommunity.microsoft.com/t5/azure-ai-news-and-announcements/ai-phi-3-mini-4k-instruct-gguf-ide-to/ba-p/4121315)

[azure.microsoft.com/en-us/blog ... -possible-with-slms/](https://azure.microsoft.com/en-us/blog/phi-3-mini-4k-instruct-gguf-possible-with-slms/)

© 2024 Science X Network

Citation: Microsoft claims that small, localized language models can be powerful as well (2024, April 24) retrieved 6 May 2024 from <https://techxplore.com/news/2024-04-microsoft-small-localized-language-powerful.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.