Why Meta is allowing users to see the inner workings of its new AI chatbot

August 10 2023, by Yali Du

Credit: AI-generated image (disclaimer)

The AI division of Mark Zuckerberg's Meta recently unveiled its Llama 2 chatbot. Microsoft has been appointed as Meta's preferred partner on Llama 2, which will be available through the Windows operating system.

Meta's approach with Llama 2 contrasts with that of the company

OpenAI, which created the AI chatbot ChatGPT. That's because Meta has made its product open source—meaning that the original code is freely available, allowing it to be researched and modified.

This strategy has sparked a vast wave of discussions. Will it foster greater public scrutiny and regulation of large language models (LLMs)—the technology that underlies AI chatbots such as Llama 2 and ChatGPT? Could it inadvertently empower criminals to use the technology to help them carry out phishing attacks or develop malware? And could the move help Meta gain an advantage over OpenAI and Google in this fast-moving field?

Whatever happens, this strategic move looks set to reshape the current landscape of generative AI. In February 2023, Meta released its first version of the LLM, called Llama, but made it available for academic use only. Its updated version, Llama 2, features improved performance and is more suitable for business use.

Like other AI chatbots, Llama 2 had to be trained using online data. Exposure to this vast resource of information helps it improve what it does—providing users with useful responses to their questions.

An initial version of Llama 2 was created through "supervised fine-tuning", a technique that uses high-quality question-and-answer data to calibrate it for use by the public. It was further refined with human feedback reinforcement learning which, as the name suggests, incorporates people's assessments of the AI's performance to align it with human preferences.

Guaranteed benefits

Meta's embrace of the open-source ethos with Llama 2 allows it to capitalize on what appears to be an approach that has worked for the

company in the past. Meta's engineers are known for their development of products to aid developers such as React and PyTorch. Both are open source and have become the industry standard. Through them, Meta has set a precedent of innovation through collaboration.

The release of Llama 2 holds the promise of safer generative AI. Through shared wisdom and collective exploration, users can identify erroneous information and any vulnerabilities that could be exploited by criminals. Unexpected applications have already emerged, such as a version of Llama 2 that can be installed on iPhones and was created by users, underscoring the potential for creativity within this community.

But there are limits to how far Meta will allow Llama 2 users to commercialize its AI system. If any party achieves more than 700 million active users in the preceding calendar month for a product based on Llama 2, it will have to request a license from Meta. For Meta, this opens up the potential for profit-sharing on successful products based on Llama 2.

Meta's strategy contrasts starkly with the more guarded approach of its primary competitor, OpenAI. Even as some question Meta's ability to compete in this area and commercialize products as OpenAI has done with ChatGPT, Meta's decision to invite worldwide developers into the fold suggests a broader vision. It's a move that positions Zuckerberg's company not merely as a player but a facilitator, harnessing global talent to contribute to the growing ecosystem of Llama 2.

This strategy could also be a shrewd hedge against potential competition from fellow tech giants such as Google. With a large population of users exploring the potential of Llama 2, any successful advance can be promptly integrated into Meta's other products. Only time will reveal the full impact of this decision, but the immediate effects on the industry are already resonating far and wide.

Advantages and pitfalls for users

The public experimentation aspect of open source technology allows for greater scrutiny, providing an opportunity for a community of users to assess Llama 2's strengths and weaknesses, including its vulnerability to attacks. The public's watchful eye may reveal flaws in LLMs, prompting the development of defenses against them.

On the downside, concerns have emerged that this is akin to "handing a knife to criminals", as it could also allow malicious users to exploit the technology. For example, its power could help fraudsters build a dialog system that generates plausible automated conversations for telephone scams. This potential for misuse has led some to call for regulation of the technology.

But exactly what rules are devised, who gets the power to supervise this process, and exactly what needs more or less scrutiny, all require careful planning to make sure that regulation does not simply prop up monopolies for the big tech companies.

In the evolving saga of AI development, the debate over open sourcing serves as a reminder that technological advancements are rarely simple or one-dimensional. The implications of Meta's decision are likely to ripple across the tech world for years to come. While Llama 2 may not yet rival the capabilities of ChatGPT, it opens the door to the development of a host of innovative products.

Google will also be under scrutiny, as speculation grows about how it may respond. In an era where open source culture thrives, it would not be surprising to see Google follow suit with its own releases.

The term "tech for good" has become a common mantra to describe technology companies using some of their resources to make a positive

impact on all our lives. Ultimately, though, this objective remains a shared responsibility, not just something that a handful of companies should be engaged in.

It's also an aim that demands collaboration and a concerted effort across academia, industry, and beyond. As LLM technologies continue to evolve, the stakes are high, and the path forward is laden with both opportunities and challenges.

Provided by The Conversation