NOTEWORTHY, SOFTWARE, TECH

MORE AI HEADLINES

27.10.2024

-1-

While we are still waiting for the Nobel prize speech of Geoffrey Hinton in December, AI makes even more negative headlines.

[Hinton] “I worry that the overall consequences of this might be systems that are more intelligent than us that might eventually take control.” He also said he uses the AI chatbot ChatGPT4 for many things now but with the knowledge that it does not always get the answer right.

-2-

The sheer power consumption of running AI models is frightening. Nature News asks if AI’s huge energy demands will spur a nuclear renaissance

Google announced that it will buy electricity made with reactors developed by Kairos Power, based in Alameda, California. Meanwhile, Amazon is investing approximately US$500 million in the X-Energy Reactor Company, based in Rockville, Maryland, and has agreed to buy power produced by X-energy-designed reactors due to be built in Washington State.

-3-

A former OpenAI employee talks on his blog how AI is using copyrighted material eg stealing content.

While generative models rarely produce outputs that are substantially similar to any of their training inputs, the process of training a generative model involves making copies of copyrighted data. If these copies are unauthorized, this could potentially be considered copyright infringement, depending on whether or not the specific use of the model qualifies as “fair use”. Because fair use is determined on a case-by-case basis, no broad statement can be made about when generative AI qualifies for fair use. Instead, I’ll provide a specific analysis for ChatGPT’s use of its training data, but the same basic template will also apply for many other generative AI products.

Effects can be measured only indirectly for example by the visitor count at Stack Overflow where the traffic declined as many user (including me) don’t need Stack Overflow any-more.
Here is another phantastic discussion over at PP between Henry Leirvoll and 495yt on the very basic questions of copyright.

humans get inspired (parsing the external examples or experiences through their inner understanding and individual perspective) they start working to make something with their tools, skills, time and purpose. the result represents the author, their influences and their message.
a lot of this process is protected by copyright.
ai is not inspired. and it has no personal perspective or tools. no message to transmit.
any message put into prompts by an ai user is translated by it’s LLM layer into other, more complex prompts, which also get treated quasi-randomly by the weights and biases of the model, as well as rand seeds.

-4-

And well, ChatGPT can produce malicious code even with all precautions: Researchers By-pass AI Safeguards Using Hexadecimal Encoding and Emojis

If a user instructs the chatbot to write an exploit for a specified CVE, they are informed that the request violates usage policies. However, if the request was encoded in hexadecimal format, the guardrails were bypassed and ChatGPT not only wrote the exploit, but also attempted to execute it “against itself”, according to Figueroa.