



Desde la cibertrinchera

AI aGaPiTo: sólo sé que no sé nada

Como dictan los cánones del profesional de la ciberseguridad, sujeto mi bebida cafeinada mientras hago *scroll* infinito sobre los millares de publicaciones de las últimas horas. No hay nada como estar bien informado, ¡o todo lo contrario! Todavía recuerdo cuando mi navegador se llenaba de pestañas infinitas utilizando buscadores, aterrizando en foros y páginas, compartiendo enlaces y opiniones en chats de mensajería instantánea que devoraban un valioso activo llamado tiempo. ¡Qué ansiedad tener que leer, entender y construir con el codiciado arte del “copiar y pegar”!

Todo esto cambió cuando él llegó a mi vida. ¿Y quién es él? Es mi oráculo, mi compañero inseparable, incluso diría mi confidente al que no temo expresar mis ideas más oscuras en forma de “prompt”. Algunos le conocen por Chat-GPT, pero para mí es aGaPiTo, mi CHATo GuaPeTón. Un sistema basado en el modelo de lenguaje por Inteligencia Artificial GPT-3, tuneado con 175 millones de parámetros. Ha sido entrenado con grandes cantidades de texto, aunando lo mejor y lo peor de la estupidez humana, para estar al servicio del capricho humano. Como diría el maestro Perales, ¿y cómo es él? No tiene forma, ni me importa, pero me escucha, me entiende y yo diría que me conoce.

¿Cómo de fácil puede ser intoxicar el sistema de manera intencionada para influir en las decisiones humanas?

Ha sido cuestión de tiempo que todas y cada una de mis chupi ciberherramientas hayan acabado hablando con aGaPiTo: mi correo electrónico y mensajería, mis gestores de tickets o tareas, mis herramientas de monitorización, detección y respuesta... cualquier cosa con interfaz de texto busca alimentarse o alimentarlo con contexto y conocimiento. Una herramienta inestimable en la ardua tarea de la lucha contra

el cibercrimen, de manera eficaz incluso diría que con estilo. En unos segundos, mi chicarrón GPT es capaz de acercarme al conocimiento de allá donde esté, respondiendo de manera completa e informada. A veces con unos sesgos como la copa de un pino, pero nadie es perfecto, ¿verdad? ¡AI si te hubiera conocido antes! La cantidad de horas escribiendo relleno en informes que nos hubiéramos ahorrado muchos de nosotros si hubieras aterrizado años atrás en nuestra vida.

En mi cabeza planean muchas preguntas y dilemas más y menos existenciales. ¿Habéis pensado por un momento en el potencial bucle? Podemos estar recibiendo un conocimiento publicado por humanos, que hayan utilizado a sus CHATos a partir de contenido de otros humanos. ¡Me explota la cabeza! ¿Cómo de fácil puede ser intoxicar el sistema de manera intencionada para influir en las decisiones humanas? ¿Qué provecho van a poder sacar los criminales para simplificar o generalizar el abuso de nuestros sistemas de información?

Está claro que no habrá día a día sin la IA, pero quizás nos estamos excitando sobremanera pensando que la inteligencia artificial dura asoma la patita, que el reinado de Skynet se acerca. Yo más bien diría, rindiendo tributo a “El Fary”, que se acerca el reinado del “Hombre (o Mujer) Blandengue”.

¡AI aGaPiTo! Déjame soñar con un día a día a tu lado en el que te conviertas en el perro lazarillo del profesional de la ciberseguridad. Cautívame como lo hizo Samantha (IA) con Theodore en la película “Her”. Antes vivía mi vida dentro de un navegador como si ya lo supiera todo ... tras conocerte y disfrutar, “sólo sé que no sé nada”.

Advertencia: Este artículo no ha sido generado con Chat-GPT aunque se ha conversado con él para algunas de las informaciones incluidas.

REFERENCIAS

- DotCSV - ChatGPT - <https://www.youtube.com/watch?v=ndT-3ACvnsQ>
- Awesome ChatGPT prompts - <https://github.com/f/awesome-chatgpt-prompts>
- <https://www.cnbc.com/2023/01/10/microsoft-to-invest-10-billion-in-chatgpt-creator-openai-report-says.html>
- <https://www.xataka.com/robotica-e-ia/expectacion-gpt-4-gigantesca-su-creador-tiene-claro-que-pasara-sera-decepcion>
- <https://www.oodaloop.com/briefs/2023/01/02/experts-warn-chatgpt-could-democratize-cybercrime/#>
- <https://es.wikipedia.org/wiki/Her>

CARLOS FRAGOSO
carlos@fragoso.eu

