La IA y Tu Privacidad

Santiago Luis Rivas Betancourt
SIA - ITBA

June 21, 2024

1 Introducción

Con el advenimiento de las nuevas tecnologías de Inteligencia Artificial y su acceso abierto a todos, muchas personas comenzaron a utilizarlas en su día a día. En particular, muchas personas lo incorporaron a su vida personal como laboral, utilizando los beneficios de los LLMs para realizar su trabajo. Esto trae varios conflictos relacionados a la privacidad de los datos no solo en el ámbito personal, pero también hay repercusiones en el ecosistema empresarial. En empresas como Apple, Wired, Verizon Communications se han introducido restricciones a la hora de que sus empleados utilicen tecnologías como ChatGPT y otras Inteligencias Artificiales[1]. Estas medidas se tomaron con diferentes objetivos, una de ellas siendo, el resguardamiento de información privada.

En particular, muchos programadores comenzaron a utilizarlo como complemento de su educación y en su trabajo. Desde generación hasta corrección de código, los LLMs han tenido éxito en una gran variedad de aplicaciones en este rubro. Con Github Copilot y ChatGPT un programador puede generar grandes cantidades de código, como también buscar optimizar o mejorar código ya existente. Si bien se podría discutir que tan útiles son estas herramientas, y que tan buena es la calidad de código generado por estos LLMs, este no sera el enfoque que se tomara. Como todo instrumento, la Inteligencia Artificial puede utilizarse de manera correcta como incorrecta y a algunos podrá servirles en mayor o menor medida para su flujo de trabajo. Más bien, el enfoque de este artículo sera el de como un programador o un equipo de programadores puede incorporar la Inteligencia Artificial en su flujo de trabajo de una manera privada y resguardando datos personales, y los de la empresa en la que trabaja.

2 El Problema

En esta era de redes sociales y de aplicaciones web, las personas ya no piensan tanto en como resguardar sus datos personales. Muchas personas entregan libremente sus datos a empresas como Meta, Google, Microsoft y ahora OpenAI. La mayoría de las personas, incluyendo programadores, no le dan importancia a la privacidad de su información. Desde publicaciones sobre viajes personales hasta utilización de sistemas cloud para resguardar documentos en la nube como Google Drive, uno termina confiando en la seguridad del sistema en el cual esta publicando dicha información. Todos confían plenamente en que estas empresas utilizan buenas practicas de encriptación y protocolos seguros de comunicación para mantener sus datos a salvo. Además, pretenden que estas empresas no venden dicha información o que si lo hacen es de manera "anónima".

Las empresas también enfrentan este problema desde otro punto de vista. La divulgación de información privada de la empresa por medio del Internet es un obstáculo serio que toda corporación debe abordar. Es normal que una empresa controle que es lo que sus empleados divulgan por sus redes sociales u otras plataformas. En efecto, no seria beneficioso para una empresa que sus secretos caigan en manos de otra empresa de tecnología como OpenAI. Un ejemplo de esto ocurrió en Samsung, donde dicha empresa permitió el uso de ChatGPT por parte de sus empleados, los cuales al utilizarlo, enviaron información confidencial (como el código fuente de un nuevo programa) a ChatGPT[2]. Hoy ChatGPT tiene la posibilidad de tener chats "secretos"[3], pero esto sigue siendo un riesgo para una empresa, ya que se debe confiar en esta promesa que OpenAI le hace al mundo. Nadie puede realmente garantizar la verdad de esta promesa, del mismo modo que nadie puede verificar la seguridad de las practicas de manipulación de datos que realiza cualquier plataforma online. Software que no corre en la computadora de uno, es software no confiable.

Estos problemas ya existían, y con la aparición de ChatGPT, las personas no piensan dos veces en compartir sus preferencias, gustos, información personal y laboral. Cada uno puede evaluar por si mismo que datos personales le ha entregado a OpenAI por medio del gran secretario GPT. Pero para muchos parece no haber alternativa. "¿Como voy a hacer una copia de seguridad de mis fotos sin Google Drive?" "¿Como voy a editar mi documento de texto sin Microsoft Word?" "¿Como voy a usar Inteligencia Artificial sin ChatGPT?". Estas son todas preguntas que alguien que realmente quiere despegarse del mundo del software corporativo puede responder. Muchas veces requiere más esfuerzo, pero es cuestión de cuanto uno valora su privacidad.

3 La Solución

Las empresas grandes encontraron la manera de resolver esto, creando sus propios modelos[1]. En Google está Bard y en Microsoft está Copilot. ¿Pero que queda para los emprendimientos pequeños o los programadores independientes? El mismo problema que enfrentan las empresas grandes, estos también lo tienen. Para una start-up, con una idea revolucionaria e innovadora, podría correr el riesgo de estar divulgando secretos que podrían llevar al éxito del emprendimiento. Archivos sensibles de configuración, pueden contener contraseñas o llaves para el acceso de servicios críticos. Para programadores, que quieren utilizar dichas tecnologías, pero son responsables con sus datos y los de su empleador, también seria beneficioso resolver este problema.

Frente a este problema, una posible solución es el self-host. Frente a la problemática que se esta analizando, el self-hosting puede resolver el problema presentado y además traer beneficios para los usuarios. Para self-hosting de Inteligencia Artificial existen varias alternativas entre ellas, Ollama[4] y TabbyML[5], plataformas que permiten el uso de LLMs que corren dentro del sistema del usuario. TabbyML se enfoca en el área de completar código, pero además provee una interfaz gráfica que permite el análisis de métricas realizados por un equipo de programadores. Ollama maneja la descarga y ejecución de diferentes LLMs predeterminados (Llama o Mistral) o hasta crear nuevos a partir de modelos en *huggingface.co*. Por su parte, *huggingface.co* es una plataforma que fomenta el desarrollo de IA fuente abierta que es competitivo con los productos de OpenAI y otras grandes empresas. También provee interfaces pare el entrenamiento y fine tuning de diferentes modelos. Además, estas soluciones están apuntadas al uso de hardware comercial y accesible.

4 Ventajas y Desventajas

El primer beneficio es que resuelve el problema sobre el resguardamiento de datos. Todo lo introducido a estos modelos, quedara en el sistema local que este corriendo Ollama o TabbyML. Esto asegura que el código o las credenciales que pueden ser leídas por un analizador de código o por un chat de IA, no estén bajo la custodia de terceros.

Un segundo beneficio es el fine tuning. Si bien es posible realizar fine tuning para ChatGPT, esto tiene un costo adicional. Teniendo un sistema local, este podría realizarse de manera independiente, y es cuestión de invertir el tiempo y recursos necesarios para realizar el fine tuning de manera local. Además, si el fine tuning se quiere realizar sobre un dataset o codebase privado, esta solución seria más adecuada.

Finalmente, como estos modelos pueden correr de manera local, la inversión puede llegar a ser menor. Si la potencia requerida para la ejecución de estos modelos es capaz de ser sostenida por el hardware que los programadores ya utilizan, entonces la inversión es nula. En caso de querer un sistema con más potencia, tal vez requiriendo comprar o alquilar hardware adicional, sera cuestión de un análisis de costos[6][7]. Para un programador independiente puede llegar a ser una limitante debido a no tiene los recursos que podría manejar una empresa. Pero para contrarrestar este punto, hoy en día, el hardware comercial ya esta al nivel requerido para correr estos modelos, más aun si se habla de productos apuntados a programadores que tienden a tener mayor capacidad que el hardware comercial promedio.

5 Conclusión

El ecosistema de código abierto y orientado a la privacidad de datos es uno que siempre tuvo su lugar para diferentes casos de uso. En el marco de la Inteligencia Artificial y con la evolución del hardware, es cada vez más tangible el poder utilizar estas tecnologías de manera local y privada. Para empresas de todo tipo, como para programadores independientes, ya existe esta posibilidad y debe ser tomada en cuenta para el desarrollo y evaluación de un proyecto. Dicho esto, también debe ser un llamado para orientar la creación de software basado en IA que este orientado a la privacidad de los datos. Esto ultimo incluye a la privacidad a nivel empresarial, pero también la privacidad de los datos de los consumidores finales que no necesariamente tienen el conocimiento para comprender que es lo que ocurre con la información que le proveen a las grandes empresas de tecnología por medio de sus servicios.

References

[1] Jonathan Gillham, "Company AI Policy Examples and Templates - Who Has Banned ChatGPT" *originality ai*, [Online]. Available: https://originality.ai/blog/ai-policy. [Accessed: 2024-06-19].

[2] Lewis Maddison, "Samsung workers made a major error by using ChatGPT" *Techradar*, [Online]. Available: https://www.techradar.com/news/samsung-workers-leaked-company-secrets-by-using-chatgpt. [Accessed: 2024-06-19].

[3] OpenAI, "New ways to manage your data in ChatGPT" *OpenAI*, [Online]. Available: https://openai.com/index/new-ways-to-manage-your-data-in-chatgpt. [Accessed: 2024-06-19].

[4] Ollama *Ollama*, [Online]. Available: https://www.ollama.com/. [Accessed: 2024-06-19].

[5] TabbyML *TabbyML*, [Online]. Available: https://tabby.tabbyml.com/. [Accessed: 2024-06-19].

[6] Rucy, "Self-hosted LLMs: Are they worth it?" *Medium*, [Online]. Available: https://medium.com/pipedrive-engineering/self-hosted-llms-are-they-worth-it-1676cbeb4f31. [Accessed: 2024-06-19].

[7] Michiel De Koninck, "Should we fine-tune a LLM for this use case? Or consider other techniques?" *ML6*, [Online]. Available: https://www.ml6.eu/blogpost/fine-tuning-large-language-models. [Accessed: 2024-06-19].