OPINIÓN

La destrucción del internet público y abierto por la Inteligencia Artificial

Por Ernesto Ángeles · 22 de septiembre de 2023 · 7 min de lectura



¿Has notado que últimamente varias apps, sitios web y plataformas están actualizando sus términos y condiciones? ¿Te has preguntado a qué se debe, para qué o siquiera, qué implica esto? ¿podría afectarte de algún modo?

Si usas los productos y servicios digitales de empresas como Google, Meta, Mircrosoft, Twitter/X , Zoom, entre otros, es muy probable que en los últimos meses te hayas encontrado con alguna notificación pidiéndote que aceptes una actualización en los términos y condiciones del servicio, los cuales muy probablemente aceptaste por el simple hecho que si no los aceptas, no puedes acceder a sus productos.

La razón de tal avalancha de nuevos términos y condiciones es que actualmente estamos en medio de un proceso de ajuste en la economía digital, en donde las empresas están recalibrando sus reglas con el fin de limitar la disponibilidad de datos en el internet público, así como la monetización de su acceso; en traducción, estamos en medio de un proceso de privatización de lo poco público que quedaba de internet.

Este proceso de cambio viene acompañado de una serie de medidas que las empresas están tomando, ya sea en la creación de barreras técnicas,el cobro por suscripción o el fortalecimiento de la capacidad de para recabar y administrar datos personales. Obviamente estos cambio significan un reajuste en la relación de las empresas con los clientes y otros actores interesados, es por esto que las empresas han echado mano de los términos y condiciones para re-calibrar las reglas del juego.

Para comenzar, los términos y condiciones son reglas, pautas y acuerdos legales que una empresa o plataforma en línea establece para regular la relación entre la empresa y los usuarios que utilizan sus servicios digitales; entre las áreas que abarcan los términos y condiciones están: las responsabilidades del usuario y de la empresa, la privacidad y la recopilación de datos, las políticas de uso aceptable, las restricciones de contenido, las reglas de propiedad intelectual y otros aspectos legales relevantes.

La relación que las empresas establecen con los usuarios por medio de los términos y condiciones usualmente es bastante desigual y desproporcionada, no sólo en su carácter de obligatoriedad para usar y acceder a un servicio que hoy en día se pensaría esencial (en algunos contextos), sino porque la extensión, lo enmarañado y ambiguo de algunos textos legales sirve como herramienta para ofuscar y confundir a quienes lo lean.

Y uno se preguntará, ¿qué ha motivado a las empresas a cambiar las reglas o por qué creen necesario el reajustar los modelos de negocio? La razón de tal cambio se debe a la inteligencia artificial y la explosión de productos de los que estamos siendo testigos; ya que los adelantos técnicos de esta tecnología han permitido escalar los modelos de inteligencia artificial a actividades más amplias, así como la capacidad de generar contenido (audio, imágenes, texto o video). Sin embargo, tal fuerza de creación y transformación viene con un costo: la necesidad de inmensas cantidades de datos para entrenar los modelos de inteligencia artificial, ya que el contenido que genera la inteligencia artificial no se crea de la nada, sino que depende de los datos.

En el proceso de creación y alimentación de modelos de inteligencia artificial diversas empresas y otros actores han tenido que echar mano de una serie de estrategias para hacerse de grandes cantidades de datos, tal como comprar el acceso a ellos, recabar sus propios datos o, en su defecto, aprovecharse de la apertura de internet y la web y hacerse de los datos de terceros actores por medio de una técnica conocida como “scrapping web” o raspado web, la cual es una técnica que permite extraer información de sitios web de manera automatizada.

El problema es que los entornos en donde se encuentran resguardados los datos no estaban preparados para tal voracidad de datos, ni mucho menos para los problemas éticos, morales, técnicos y de propiedad intelectual que esto significa; ya que el raspado de la web no distingue entre datos personales, propiedad intelectual o información sensible, por lo que el carácter “público” de internet se está volviendo un dolor de cabeza para quienes generan y resguardan datos, mientras que resulta una mina de oro para quienes desarrollan modelos de inteligencia artificial.

Tan sólo pensemos en el caso de Wikipedia, una enciclopedia digital colaborativa y de acceso gratuito, cuyo enorme acervo de datos puede ser apropiado por una empresa para desarrollar un modelo de inteligencia artificial privado, lo que en la práctica significa que una empresa puede adueñarse del trabajo de los demás sin un pago de por medio y ganar dinero en el proceso.

Esto ha generado dos movimientos bastante desiguales: por una parte las empresas, gremios y demás grupos organizados están buscando evitar que las empresas de tecnología sigan robando sus datos, por lo que no sólo están erigiendo restricciones más elevadas, sino que se han sumado a la carrera de convertirse en grandes recolectores y cuidadores de datos, lo que en términos prácticos implica aún más controles que afectarán a los usuarios, tal como tarifas de pago, restricciones de código y otras acciones que se traducen en una mayor división digital, en donde el acceso a la información será más restringido y opaco, sin mencionar la falta de transparencia.

Pensemos en el caso de Twitter o 4Chan, dos de las redes sociales que más se caracterizaban por su transparencia y apertura, las cuales en los últimos meses han anunciado una serie de restricciones en su acceso a datos, especialmente por medio del cobro por acceso y la creación de barreras técnicas para evitar o dificultar el raspado web.

También está el caso de las empresas que están creando barreras legales; por ejemplo el periódico New York Times, el cual el cual, por medio de actualizar sus términos y condiciones, prohibió que su contenido se use en el desarrollo de “cualquier programa de software, incluido, entre otros, el entrenamiento de un sistema de aprendizaje automático o inteligencia artificial”.

Por otro lado, las empresas están reajustando las reglas con los usuarios, especialmente en lo que se refiere al recabo de datos, esto con el fin de evitar cualquier problema que pueda detener la voracidad de las empresas por los datos y su uso en el desarrollo y entrenamiento de modelos de inteligencia artificial, por lo que ya no sólo se recabarán tus datos para saber más de ti y venderle ese conocimiento a las empresas de mercadotecnia y publicidad, sino que ahora también se usarán para crear productos basados en inteligencia artificial, una ganancia doble de la cual los usuarios no obtendrán nada en absoluto, sólo el mismo proceso de creación y optimización de servicios, varios de los cuales son o serán de cobro.

Las empresas que han actualizado sus términos y condiciones para que los usuarios les permitan usar sus datos en el entrenamiento de modelos de inteligencia artificial son: Microsoft, Google,Twitter, Meta (incluye Facebook, Instagram, Messenger y sus productos de inteligencia artificial), así como Zoom; aunque de todas estas, el único cambio que generó suspicacia y rechazo por parte de los usuarios fue Zoom, mientras que las otras empresas han gozado del silencio de la comunidad y la complicidad en el uso y aceptación de los términos y condiciones.

En este punto queda preguntarnos: Más allá del usual abuso que cometen las empresas en contra de los usuarios, ¿Cómo puede afectar esto al futuro del supuesto internet abierto? Y ¿Cuál será su impacto en el desarrollo y actualización de servicios y productos de inteligencia artificial?

Por último, queda pendiente el analizar la necesidad de reformar en México a la institución encargada de la protección de datos, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI), ya que, dada la naturaleza de sus funciones (la protección de los datos personales y el acceso a la información pública), el organismo se ve rebasado por el rumbo que están tomando la relevancia estratégica de los datos en general, debido sobre todo al ascenso de la tecnología conocida como inteligencia artificial generativa. Ya ni que decir sobre su deficiente labor en la protección de datos personales antes de esta coyuntura o su complicidad con el mercado de datos y en detrimento de los usuarios y los mexicanos.

Publicado en Punto de Referencia. Dirección original: https://puntodereferencia.mx/opinion/la-destruccion-del-internet-publico-y-abierto-por-la-inteligencia-artificial
Descargado el 21 de noviembre de 2026.