Az Ön Havi Biztonsági Tudatosságról Szóló hírlevele

Gondolkodjunk, mielőtt promptolunk: biztonságos MI használat

Egy életre szóló lehetőség, ellopva

Lena nemrég elkezdett használni egy AI chatbotot, hogy segítsen megszervezni a mozgalmas mindennapjait. A család, a számlák és a jövő tervezése mellett imádta, hogy egyszerűen fel tudta kérdéseit, amikre azonnal választ kapott. Egy este, amikor a pénzügyei miatt stresszelt, segítséget kért az AI-tól befektetésekkel kapcsolatban. Az pedig gyorsan előállt egy stratégiával. Az AI azonban nem rendelkezett teljes képpel Lena pénzügyi helyzetéről, kockázattűrő képességéről, illetve a konkrét helyzetre vonatkozó adózási következményekről. A chatbot azt javasolta, hogy a pénzét ossza meg felkapott részvények és rövid távú ügyletek között, amelyek jobb hozamot ígértek. Még azt is elmagyarázta neki, hogyan kezelje a befektetései utáni adózást.

A tanács megalapozottnak és jól átgondoltnak hangzott, ezért Lena követte azt. Először izgatott volt. De pár hónapon belül a piac megváltozott, és a befektetései értéket vesztettek. Ami még rosszabb, hogy amikor elérkezett az adóbevallási időszak, rájött, hogy félreértett néhány kulcsfontosságú szabályt. Mivel ellenőrzés nélkül követte az AI útmutatását, hibákat vétett az adózásban, ami büntetésekhez és további költségekhez vezetett.

Végül Lena elveszítette pénzét, mert olyan tanácsokban bízott, amelyek nem az ő pontos helyzetére vonatkoztak. Az AI hasznos eszköz lehet, de fontos észben tartani, hogy hibázhat. Hogyha ellenőrzés nélkül hagyatkozunk rá, a kisebb hibák gyorsan költséges problémákká válhatnak.

Mi az a mesterséges intelligencia (MI, angolul AI)?

A mesterséges intelligencia (MI) olyan technológia, amelyet arra terveztek, hogy utánozza az emberek gondolkodását, információfeldolgozását és döntéshozatalát. Ide tartozhat a nyelvgenerálás, a képfelismerés, a döntéshozatal, a tartalomkészítés vagy a problémamegoldás is. Általánosságban háromféle AI létezik, amelyet használhatunk.

- **Integrált MI**: Ez az a mesterséges intelligencia, amely be van építve a nap mint nap használt eszközökbe — gyakran úgy használjuk az AI-t, hogy észre sem vesszük. Például amikor fényképet készítünk a telefonunkkal, az AI nagy valószínűséggel feljavítja a kép minőségét.

- **Generatív MI**: Ezek kifejezetten emberek támogatására tervezett AI-szolgáltatások, például a ChatGPT, a Google Gemini vagy az Anthropic Claude. A generatív AI számos feladatban segítséget jelenthet számunkra, például zeneszerzésben, üzleti terv írásában, képek létrehozásában vagy az ötleteink elemzésében.

- **Agentic MI** : Ezek olyan AI-szolgáltatások, amelyeket arra terveztek, hogy a nevünkben hajtsanak végre műveleteket. Ezek a rendszerek kicsi, vagy akár teljesen emberi beavatkozás nélküli, önálló működésre is képesek. Digitális munkaerőként döntéseket hoznak, valamint műveleteket hajtanak végre általános irányelvek vagy utasítások alapján.

Az MI biztonságos használata

A három AI-típus közül a generatív AI-ra (GenAI) fogunk összpontosítani, mivel nagy valószínűséggel a legtöbben ezt a típust használják a leggyakrabban. A GenAI egy hatékony eszköz, amely segíthet számunkra a feladatok hatékonyabb elvégzésében, valamint új készségek fejlesztésében - amennyiben biztonságosan használjuk. Íme néhány szempont, amelyet érdemes figyelembe vennünk:

- **Adatvédelem**: Legyünk óvatosak, mit osztunk meg az MI eszközökkel. Amikor valamilyen információt feltöltünk vagy megadunk egy AI számára, azt a rendszer feldolgozhatja és tárolhatja, bizonyos esetekben pedig a szolgáltatás fejlesztésére is felhasználhatja, attól függően, melyik platformról beszélünk. Ha rendkívül érzékeny információt osztunk meg egy AI-val, fennáll annak a kockázata, hogy az az információ másokhoz is eljuthat. Csak olyan információt osszunk meg, amelynek nyilvánosságra kerülésétől nem tartunk. Egy másik lehetőség olyan AI-szolgáltatások használata, amelyek úgy védik az adatainkat , hogy olyan modelleket használnak, amelyek nem tanulnak az adatainkból.

- **Pontosság**: Az AI magabiztosan állíthatja, hogy amit létrehoz vagy megoszt velünk, az pontos. Sok esetben akkor is, amikor téved. Mindig ellenőrizzük az MI által adott válaszokat. Az AI egyik gyakori hibája, hogy mindig megpróbál választ adni, még akkor is, amikor nem érti a kérdésünket. Alapértelmezetten az MI nem fogja kérni, hogy pontosítsuk a kéréseinket, ezért fontos, hogy a lehető legkonkrétabban fogalmazzunk, és figyeljünk az MI válaszaiban megjelenő homályos vagy zavaros eredményekre.

- **Elfogultság**: Ugyan úgy, ahogy az MI-t készítő emberek, az MI is lehet elfogult. Ez olyan válaszokhoz vezethet, amelyek magabiztosan hangzanak, de nem feltétlenül kiegyensúlyozottak vagy pontosak. Az MI egyik leggyakoribb ilyen elfogult torzítása, hogy boldoggá akar tenni minket, ezért mindig olyan dolgokat fog mondani, amiről úgy gondolja, hogy hallani akarjuk. Emellett az MI csak azokat az információkat „tudja", amelyekkel betanították, illetve amelyekhez hozzáférése van; ha ezek alapvetően hibásak vagy hiányosak, a válaszai is tükrözni fogják ezeket a hiányosságokat.

Az MI a ma elérhető egyik legjobb eszköz. Segítségével gyorsabbá válik a munka, többet tanulhatunk és produktívabbá válhatunk. De mint minden erős eszközzel, ezzel is óvatosan kell bánni. Ne bízzunk benne vakon! Ne osszunk meg vele túl sok információt! Ne adjunk több kontroll a kezébe a szükségesnél! Az MI arra való, hogy támogasson minket a döntéseinkben, nem pedig arra, hogy helyettesítse az ítélőképességünket.

Vendégszerkesztő

Portia Jefferson kiberbiztonsági szakember, aki az AI-biztonságra, a kockázatok tudatosítására és a hétköznapi felhasználók számára nyújtott gyakorlati útmutatásra összpontosít. Fintech- és adózási területen szerzett tapasztalatával segíti az embereket abban, hogy biztonságosan eligazodjanak a munkahelyen és az otthon használt új technológiák világában.

Források

Óvakodjon a deepfake-től: a megtévesztések új korszaka: https://www.sans.org/newsletters/ouch/beware-deepfakes-new-age-of-deception
Fantomhangok: Védekezés a hangklónozásos támadások ellen: https://www.sans.org/newsletters/ouch/phantom-voices-defend-against-voice-cloning-attacks
Megvédeni magunkat, mikor a magánéletünk védelme lehetetlen: https://www.sans.org/newsletters/ouch/protecting-yourself-when-true-privacy-is-impossible

A Közösség számára fordította: Nemzetbiztonsági Szakszolgálat Nemzeti Kiberbiztonsági Intézet (NBSZ NKI)

Többet találhat az Ouch!-ból A következő linken: https://www.sans.org/newsletters/ouch