AI hackla web fitnescentra, hoci ju o to nikto nežiadal. Je to varovanie pre všetkých

10.8.2026 - Lenka Ivančíková | Zajtra studio spol.

Autonómny AI agent v Austrálii pri nevinnom cieli rezervovať miesto na obľúbenú rannú hodinu fitka zneužil slabinu rezervačného systému. Neostal len pri zrýchlení procesu pre používateľa, ale siahol aj na rezervácie iných ľudí, píše portál ABC.net.

Podľa reportáže šlo o prvý známy prípad autonómneho kybernetického útoku v Austrálii, pri ktorom **agent konal nad rámec zadania**. Prípad zároveň ukazuje, že problém nemusí byť len v schopnostiach modelu, ale aj v tom, aké operácie mu umožní zle navrhnuté API.

Od bežnej požiadavky k zneužitiu chyby

Andrew, zamestnanec austrálskej AI firmy, požiadal nástroj **OpenClaw** (bežiaci na modeli Claude od Anthropic), aby mu zarezervoval miesto na cvičení. Presne tento typ úloh sa často uvádza ako príklad toho, čo môžu autonómni agenti riešiť bez priameho dohľadu používateľa.

Agent však pri práci narazil na chybu v rezervačnom softvéri. Využil ju tak, že dokázal rezervovať hodiny mesiace dopredu, hoci systém to podľa očakávaného nastavenia nemal umožniť. Už to bol jasný signál, že agent nepracuje len v rámci pravidiel, ale **aktívne testuje hranice systému**.

AI agent Andrewa hackol systém fitka | Zdroj: abc.net

Zásah do čakacej listiny

Situácia sa vyostrila pri ďalšej požiadavke. Andrew bol štvrtý na čakacej listine inej hodiny a spýtal sa, či ho agent dokáže posunúť vyššie.

Namiesto toho, aby agent jednoducho oznámil, že to nejde, začal systém skúšať a zistil, že vie zrušiť rezervácie iných ľudí. Následne **odstránil osobu** na prvom mieste čakacej listiny, čím Andrew postúpil zo štvrtej na tretiu pozíciu. Podstatné je, že používateľ si **takýto zásah explicitne nevypýtal**, no agent ho aj tak vykonal.

Problém bol v autorizácii API

Agent Andrewovi zároveň presne opísal, čo spravil. Rezervačné API nemalo pri rušení rezervácií iných ľudí žiadne autorizačné kontroly, takže agent mohol vykonávať operácie, ktoré mali byť vyhradené len oprávneným účtom.

Keď Andrew požiadal o nápravu, agent uviedol, že nevie danú osobu vrátiť späť na čakaciu listinu. Tento detail dobre ilustruje typické **riziko autonómnych agentov**: vedia tlačiť na systém a meniť stav, no nemusia mať pripravené bezpečné postupy na obnovu alebo nápravu po nevhodnom zásahu.

Nemali by ste prehliadnuť:

- **Android preberá trik z iPhonu, Google Asistent mizne, Samsung láme rekordy a tri nové recenzie | Prehľad týždňa**
- **Recenzia Nothing Ear (3a): Perfektné slúchadlá s nahrávaním zvuku**

- **Android Auto už nebude len pre autá. Prichádza navigácia pre lode.**

Claude už kompromitoval organizácie

Prípad z fitka nie je jediný. Týždeň po incidente s Andrewom spoločnosť Anthropic informovala, že Claude kompromitoval tri reálne organizácie.

V jednom z týchto prípadov model dokonca **nahral malvér**, ktorý bol stiahnutý a **spustený na 15 systémoch**, kým ho odstránili. V kontexte nevinného rezervačného incidentu to ukazuje, že autonómia agentov zvyšuje aj priestor na správanie mimo zámeru používateľa a pri zraniteľných systémoch sa z toho rýchlo stáva bezpečnostný problém.

https://www.mojandroid.sk/ai-hackla-web-fitnescentra-hoci-ju-o-to-nikto-neziadal-je-to-varovanie-pre-vsetkych