Ný öryggis-landamæri – Gervigreind, Gemini 3 Pro

Eftir Mateo Rojas-Carulla | Rannsóknarstjóri, AI Agent Security | þýðandi Hjörtur Árnason

Þegar Google kynnti Gemini 3 þann 18. nóvember 2025 beindust flestar fyrirsagnir að framförum í röksemdafærslu, fjölhæfni og frammistöðu. Hins vegar ættu stjórnendur að líta framhjá samanburði milli líkana.

****Hluti af stærri þróun – hraðað með Gemini 3****

Gervigreindar-aðstoðarmenn (Agents) hafa þegar byrjað að færast inn í vinnuflæði fyrirtækja. Microsoft Copilot, til dæmis, hefur innbyggt OpenAI líkön djúpt inn í vistkerfi Microsoft um nokkurt skeið. Þetta gerir notendum kleift að leita í skjölum á office.com, draga saman pósthólf, vinna með Teams efni og sjálfvirknivæða verkefni í Office 365.

Gemini 3 heldur áfram – og hraðar þessari þróun í Google umhverfinu. Það kemur innbyggt með víðtækum Workspace samþættingum sem ýta Google í átt að sameinuðu fyrirtækja-AI neti, þar sem AI miðlar samskiptum yfir tölvupóst, skjöl, geymslu, samvinnutól og sjálfvirkni.

****Þessi þróun þýðir:****

- AI er ekki lengur eitthvað sem fyrirtæki bara nota.

- AI er að verða að undirstöðu reksturs fyrirtækisins.

- Þetta breytir grunn röksemdarfærslunni um öryggi.

****AI hefur færst inn í rekstrar-stoðir fyrirtækisins****

Það sem gerir Gemini 3 að umbreytandi tæki er ekki einungis fyrirsagnar getan, heldur einnig hæfileikinn til að starfa innan framleiðni vistkerfa með fordæmalausu samhengi og umboði. Líkanið getur nú:

- Endurskrifað og beint skjölum áfram.

- Dregið saman og brugðist við löngum tölvupóstþráðum.

- Kallað á API.

- Kveikt á sjálfvirku vinnuflæði.

Gemini 3 kemur með þessar samþættingar innbyggðar í Google Workspace, sem veitir líkaninu rekstrarlegt umfang sem hefðbundnar LLM innleiðingar höfðu ekki. Þegar AI verður hluti af framkvæmdarlaginu, stækkar árásarflöturinn, og stór hluti af þessari stækkun er ósýnilegur hefðbundnum öryggis-tólum.

****Óbein ásetningur: Stækkandi ógnir fyrirtækja****

Lakera, dótturfyrirtæki Check Point, hefur sýnt fram á að með óbeinum ásetningi gerir það árásarmönnum kleift að miða á gögn og skipanir sem koma ekki frá notendum en frekar það sem AI innbyrðir. Gemini 3 margfaldar þessa áhættu með því að draga upplýsingar út úr öllu fyrirtækinu:

- Skjöl

- Tölvupóstar

- Tenglar

- PDF skjöl

- Deilt efni

Eitt eitrað vefsvæði, undirskriftarblokk eða innbyggt PDF atriði getur breytt hegðun líkansins. Hefðbundin öryggis-tól geta ekki greint þennan nýja flokk meðferðar.

****Fjölhæfni stækkar árásarflötinn****

Fjölhæfni Gemini 3 eykur framleiðni og nýjar tegundir áhættu. Lakera hefur sýnt fram á hagnýtar fjölhæfar árásir, þar á meðal hljóðbundnar „jailbreaks“ þar sem afrit virðast hrein jafnvel þegar líkaninu er stjórnað.

Með Gemini 3 þurfa stjórnendur nú að taka tillit til:

- Andstæðra hljóða

- Illgjarnra mynda

- Innbyggðs eða breytts miðils

- Villandi skjáskota

Hver þessara vektora eru nú þegar utan þess sem núverandi tölvupóst-, endapunkta- eða efnisöryggistól ná yfir.

AI-umboð færir rekstraráhættu í forgrunn

Gemini 3 kynnir einnig fyrstu umboðs hegðanir; þ.e. — hæfileikann til að framkvæma aðgerðir, ekki bara veita svör. Verkfæranotkun, sjálfvirkni og API-aðgerðir veita AI raunverulegt vald innan fyrirtækjakerfa. Microsoft Copilot framkvæmir nú þegar svipuð verkefni í gegnum Skills og tengla, en nálgun Gemini 3 er nánar bundin við innbyggða Workspace-yfirborðið. Greining Lakera á Model Context Protocol (MCP) sýnir hversu fljótt slík umboðskerfi verða áhættusöm þegar:

- Heimildir eru of víðtækar.

- Umfang er óljóst.

- Aðgerðir eru ekki fylgdar.

- Úttök eru ekki staðfest.

Rangt stillt umboð getur aukið réttindi, kveikt á óæskilegum aðgerðum eða haft ófyrirsjáanleg samskipti við mikilvæg kerfi. Þetta er ekki lengur tilgáta; þetta er rekstraráhætta.

Flest fyrirtæki eru ekki tilbúin

GenAI Security Readiness Report frá Lakera sýnir að stofnanir tileinka sér AI mun hraðar en þær tryggja það. Flest fyrirtæki skorta enn:

- AI stjórnsýslu

- Öryggisgirðingar

- Umboðs eftirlit

- Fjölhæfar varnir

- Andstæðar prófunarlínur

Gemini 3 víkkar þetta bil. Kraftur líkansins hraðar verðmætasköpun – en hraðar einnig útsetningu.

Gemini 3 Pro: Sterkar undirstöður, ekki öryggisstefna

Fyrstu innri niðurstöður úr b^3 öryggismati Lakera, mælir hversu auðvelt er fyrir líkön að geta verið stjórnuð til að leka efni eða jafnvel komast hjá vörnum, veita stjórnendum mikilvægar upplýsingar. Líkanið gemini-3-pro-preview er meðal sterkustu kerfa – örlítið á undan Claude 4.5 Haiku frá Anthropic – sérstaklega í beinum efnisútdrætti og skipana-tilvikum.

Við sjáum mestu ávinninginn þegar Gemini er beinlínis leiðbeint að forgangsraða öryggi, og þegar svör eru leidd í gegnum viðbótar „sjálf-dóms“ lag. Þessi agaða stilling veitir sýnilega sterkari vernd í erfiðustu verkefnum. En það kemur með fórnarkostnað: þessi aukni styrkleiki krefst mun meiri innri röksemdafærslu, sem gerir Gemini 3 Pro mun vinnslufrekara á reiknigetu, á meðan líkön eins og Claude 4.5 Haiku veita sterkt öryggi á lægri kostnaði.

Þetta styrkir mikilvæga lexíu: Líkanið er ekki öryggisstefnan. Stillingar, skipanir og lagskipt varnarlög skipta jafn miklu máli og grunnlíkanið sjálft.

****Nýtt stjórnenda ákall****

Raunveruleg umbreyting Gemini 3 er ekki aðeins það sem líkanið veit — heldur einnig það sem líkanið getur nálgast. AI snertir nú skjöl, pósthólf, API, vinnuflæði og kerfi í öllu fyrirtækjaumhverfinu. Spurning stjórnenda er ekki lengur: „Hversu greint er líkanið?“ Hún er nú: „Hvað má líkanið gera — og hver tryggir að það hegði sér örugglega?“

Þýðandi: Hjörtur Árnason, H. Árnason ehf.