

AI CYBERSECURITY SUMMIT

Kontrollált AI: Adatvédelem és költségoptimizálás on-prem AI-al és dinamikus routing-al

2026. 09. 08. 14:00

Innovation Hall

HUN · 20 perc

**Szekeres Viktor**

tulajdonos, elnök · Gloster Digital Group Nyrt.

Szekeres Viktor, tulajdonos, elnök a Gloster Digital Group Nyrt.-nél, bemutat egy megoldást az AI költségek, az adatbiztonság és a beszállítói függőség kezelésére egy központosított útválasztó rendszer segítségével.

MIRŐL VOLT SZÓ

A prezentáció a hagyományos átalánydíjas árképzéstől a határvonalas AI modellek használatalapú árképzéséhez való elmozdulást azonosítja, ami előre nem látható költségekhez vezethet. Míg az átlagos tokenköltségek csökkennek, az AI-ügynökök megjelenése azt jelenti, hogy a tokenek nagy részét már nem emberek, hanem robotok generálják. Például egy McKinsey-tanulmány szerint az AI-csapatok 93%-a túllépi tokenköltségvetését.

Ezeknek a kihívásoknak a kezelésére a vállalat olyan központi útválasztót fejlesztett ki, amely az adatérzékenység, a költségek és a belső szabályok alapján irányítja a kéréseket. Ez a rendszer lehetővé teszi, hogy az érzékeny adatokat helyi hardveren dolgozzák fel nyílt forráskódú modellek, például a QAN 3.8 használatával, míg a nem érzékeny feladatok továbbra is nagy határvonalas modelleket használhatnak. Például a rendszer sikeresen automatizálta a belső munkaidő-jelentéseket anélkül, hogy bármilyen adat elhagyta volna a helyi környezetet.

KIEMELT MEGÁLLAPÍTÁSOK

- A felhasználásalapú árképzési modellek jelentős pénzügyi kockázatokat jelentenek a vállalatok számára, mivel az AI-ügynökök egyre több feladatot automatizálnak.
- A helyi AI-telepítés olyan hardveren, mint a Spark box, kiszámítható költségstruktúrát biztosíthat és garantálhatja az adatok szuverenitását.
- A javasolt útválasztó megoldás 84–106 token/másodperc sebességet ér el, és teljesítménye összevethető a csúcskategóriás modellekével.

ÜZLETI HATÁS

A megoldás az előre nem látható felhőalapú AI-költségeket előre látható CAPEX modellé alakítja át, teljes ellenőrzést biztosítva az audit naplók és az adatvédelem felett. Integrálódik meglévő vállalati eszközökkel, például a SharePoint, Jira, SQL és SAP rendszerekkel, lehetővé téve az informatikai részlegek számára a zökkenőmentes átállást. A helyi hardver használatával a vállalatok elkerülhetik a „Shadow AI” problémát, amikor az alkalmazottak jogosulatlan eszközöket használnak, mivel a router minden felhasználó számára szabályozott, biztonságos felületet biztosít.

KÖVETKEZTETÉS

Szekeres Viktor a megoldás részleteinek megtekintésére október 7-én rendezett eseményre hívja fel a figyelmet. Ajánlásuk szerint a vállalatoknak olyan modell felé kellene elmozdulniuk, amelyben a határvonalas modellek erejét a helyi infrastruktúra biztonságával és költséghatékonyságával egyensúlyozhatják.

AI CYBERSECURITY SUMMIT

Controlled AI: Data Privacy and Cost Optimization with On-Prem AI and Dynamic Routing

2026. 09. 08. 14:00

Innovation Hall

HUN · 20 perc

**Viktor Szekeres**

owner, president · Gloster Digital Group

Viktor Szekeres, owner and president of Gloster Digital Group, argues that enterprises must transition from unpredictable cloud-based AI costs to controlled, local infrastructure to ensure data security and financial predictability.

WHAT WAS DISCUSSED

Viktor Szekeres highlights a shift in AI pricing from flat fees to usage-based models, noting that while average token costs are decreasing, total expenditures are rising due to the proliferation of autonomous agents. He points out that many AI teams currently exceed their budgets because agents generate tokens at a much higher volume than humans, for example, a McKinsey study indicates that 93% of AI teams exceed their token budgets.

To address concerns regarding high costs, vendor dependency, and regulatory compliance, the company developed a central routing solution that directs tasks to either large frontier models or local hardware. This system allows sensitive data, such as medical or HR records, to be processed on-site using models like QAN 3.8 on affordable hardware like the Spark box. These elements collectively represent a strategy to provide a sovereign, auditable, and cost-effective AI environment for large enterprises.

KEY TAKEAWAYS

- Usage-based pricing models are becoming the standard for frontier AI services.
- Autonomous agents are significantly increasing token consumption compared to human-led interactions.
- Local hardware solutions can achieve speeds of 84-106 tokens per second while keeping data behind the firewall.

BUSINESS IMPACT

The proposed solution transforms unpredictable cloud AI expenses into a predictable CAPEX model with a clear return on investment, often achievable within 5-6 months for intensive users. By integrating with existing enterprise tools like SAP, Jira, and Salesforce, the system ensures full auditability and data protection, which is critical for GDPR compliance and internal security policies.

CONCLUSION

Viktor Szekeres concludes by inviting the audience to an event on October 7 to see a detailed demonstration of the solution. He emphasizes that while AI is like electricity—available to everyone—the value lies in how a company chooses to harness it safely and efficiently.