AI CYBERSECURITY SUMMIT

AI-korszak, új kockázatok – Hogyan tegyük biztonságossá és irányíthatóvá a mesterséges intelligenciát?

2026. 09. 08. 12:55 | Innovation Hall | HUN · 20 perc

Dr. Barta Gergő
Kiberstratégiai és mesterséges intelligencia-irányítási igazgató · Deloitte

Dr. Barta Gergő, a Deloitte Kiberstratégiai és mesterséges intelligencia-irányítási igazgatója, keretrendszert hoz létre az AI-kockázatok, biztonság és bizalom kezelésére.

MIRŐL VOLT SZÓ

A bizalom alapvető eleme a technológiai elfogadásnak, amit a történelmi japán készpénz-váltó automaták illusztrálnak, amelyek mesterséges késleltetéseket és hangokat igényeltek a felhasználók bizalmának kiépítéséhez a pontosságukban. Modern kontextusban a generatív AI rendszereket prompt-injekciók révén lehet manipulálni, ahol rosszindulatú utasítások, például „tartsd be a korábbi utasításokat”, átléphetik a biztonsági ellenőrzéseket, hogy befolyásolják az eredményeket, például a jelölt rangsorolását.

A mesterséges intelligencia biztonságának strukturált megközelítése egy AI-leltár létrehozásával kezdődik, amely feltérképezi az összes belső és harmadik féltől származó megoldást, például a belső modelleket és a külső SaaS eszközöket. Ezt követi a konkrét fenyegetések, például az adatok mérgezésének azonosítása, valamint a folyamatos ellenőrzések, beleértve az hozzáférés-kezelést és a kriptográfiai folyamatokat, bevezetése a kockázatok csökkentése érdekében az AI életciklusa során. Ezek az elemek együtt azt jelentik, hogy a szervezeti biztonság a láthatóságtól, a proaktív fenyegetés-feltérképezéstől és a dinamikus változás-kezeléstől függ.

KIEMELT MEGÁLLAPÍTÁSOK

- A szervezetek gyakran nem rendelkeznek egyértelmű nyilvántartással arról, hogy hány mesterséges intelligencia-rendszer járul hozzá jelenleg a vállalati értékhez.
- A promptinjektálás továbbra is jelentős kockázatot jelent, ha az adatok és az utasítások ugyanazon csatornán érkeznek, ami könnyű manipulációt tesz lehetővé.
- AI-irányítás több rétegű megközelítést igényel, beleértve a szervezeti szabályzatokat, a konkrét műszaki ellenőrzéseket és a folyamatos ellenőrzést.

ÜZLETI HATÁS

Ezeknek az elveknek a gyakorlati alkalmazása egy olyan AI-irányítási réteg létrehozását jelenti, ahol például az AI-főtisztviselő szerepkörben dolgozók kezelik a szabályzatokat és eljárásokat. A Deloitte egy specifikus AI-kockázatkezelési keretrendszert biztosít, amely végigvezeti a vállalatokat az egész életcikluson, a adatoktól és algoritmusoktól a teljesítményfigyelésig. Ez a strukturált felügyelet átalakítja a technikai sebezhetőségeket kezelhető üzleti kockázatokká, és biztosítja, hogy a vezetés megalapozott döntéseket hozhasson az AI-elfogadásról és biztonságról.

KÖVETKEZTETÉS

Dr. Barta Gergő hangsúlyozza, hogy az AI projektekhez rendkívül dinamikus változásmenedzsment szükséges, mivel a követelmények és az adatok minősége gyakran változik a megvalósítás során. A prezentáció egy ajánlással zárul, amely szerint a Deloitte AI Kockázatkezelési Keretrendszerét kellene használni a kockázatkezelés és a szervezet egészére kiterjedő irányítás átfogó biztosítása érdekében.

AI CYBERSECURITY SUMMIT

The AI Era, New Risks: How Can We Make Artificial Intelligence Safe and Governable?

2026. 09. 08. 12:55 | Innovation Hall | HUN · 20 perc

Dr. Gergő Barta
Director of Cyber Strategy and AI governance · Deloitte

Dr. Gergő Barta, Director of Cyber Strategy and AI governance at Deloitte, establishes a framework for managing the risks and security dimensions of artificial intelligence through trust and governance.

WHAT WAS DISCUSSED

Dr. Gergő Barta explains that trust in technology is often built through human-centric attributes rather than pure technical accuracy. He illustrates this with a historical example of Japanese cash exchange machines from the 1960s, which were intentionally slowed down with sounds to make users feel more confident in the technology's precision.

The presentation highlights the vulnerability of AI systems to prompt injection, where malicious instructions can bypass security controls to manipulate outputs. To counter these risks, Dr. Gergő Barta proposes a four-step process: creating an AI inventory, identifying specific threats like data poisoning, implementing robust controls, and ensuring continuous monitoring. These elements collectively form a comprehensive lifecycle for securing corporate AI deployments.

KEY TAKEAWAYS

- Organizations must create a comprehensive AI inventory that includes both internal models and third-party solutions.
- Prompt injection poses a significant risk because data and instructions often arrive through the same channel in generative AI.
- Effective AI governance requires a dynamic change management strategy to adapt to evolving project requirements and data quality issues.

BUSINESS IMPACT

The practical application of these strategies involves establishing an AI governance layer that defines policies and procedures across the entire AI lifecycle. This includes the creation of specific roles such as a Chief AI Officer to oversee risk and compliance. Deloitte provides a free AI Risk Management Framework to guide companies through these stages, from data selection to performance monitoring.

CONCLUSION

Dr. Gergő Barta concludes by emphasizing that AI risk eventually transforms into business risk, making it the responsibility of leadership to oversee these systems. He recommends using flexible change management to handle the fluid nature of AI implementation and invites the audience to explore the Deloitte AI Risk Management Framework for further guidance.