AXION SECURITY OVERVIEW

MAGYAR MASTER — Tervezet v1

1. Biztonsági megközelítés

Az Axion többrétegű technikai és szervezési biztonsági megközelítést alkalmaz a Szolgáltatás, a felhasználói adatok, a feltöltött tartalmak és az AI-alapú munkafolyamatok védelme érdekében.

Az alkalmazott intézkedések célja többek között a jogosulatlan hozzáférés, a rosszindulatú bemenetek, az adatszivárgás, a szolgáltatással való visszaélés és az AI-rendszerek manipulációjának kockázatának megelőzése, felismerése, korlátozása és mérséklése.

Az Axion biztonsági modellje defence-in-depth megközelítésre épül: nem egyetlen ellenőrzési pontra támaszkodik, hanem több, egymást kiegészítő kontrollt alkalmazhat a felhasználói bemenettől az AI-orchestrationön és forráshozzáférésen át a végrehajtásig és az eredmények kezeléséig.

2. AI- és promptbiztonság

Az Axion biztonsági ellenőrzéseket alkalmazhat azokon a felületeken és munkafolyamatokban, ahol a felhasználók utasításokat adhatnak AI-rendszerek számára.

Az ellenőrzések többek között a prompt injection és utasításmanipuláció, rendszer- vagy belső utasítások jogosulatlan felülírására irányuló kísérletek, biztonsági vagy hozzáférési korlátozások megkerülése, AI-munkafolyamatok jogosulatlan befolyásolása, valamint nem engedélyezett adatok vagy rendszerinformációk megszerzésére irányuló kísérletek kockázatának kezelésére szolgálnak.

A biztonsági ellenőrzések az AI-munkafolyamat több pontján is alkalmazhatók. A rendszer ezért nem kizárólag a felhasználói bemenet egyszeri ellenőrzésére támaszkodik.

3. AI-végrehajtás és orchestration elkülönítése

A felhasználói utasítások nem kerülnek automatikusan közvetlen végrehajtásra. Az Axion elkülöníti egymástól a felhasználói bemenetet, az AI-orchestration folyamatokat, a források és kontextusok kiválasztását, valamint az egyes végrehajtási műveleteket.

Ez a többrétegű végrehajtási modell csökkenti annak kockázatát, hogy egy rosszindulatú vagy manipulált bemenet közvetlenül befolyásoljon privilegizált rendszerfunkciót vagy más, nem engedélyezett erőforrást.

4. Workspace-, projekt- és artifact-hozzáférések elkülönítése

Az Axion a workspace-ekhez, projektekhez és artifactokhoz kapcsolódó műveleteknél hozzáférési és scope-alapú korlátozásokat alkalmaz.

A rendszer a releváns felhasználói, projekt-, workspace- és artifact-kontextus alapján határozza meg, hogy egy adott művelet milyen erőforrásokat használhat.

Ezeknek a kontrolloknak egyik célja annak kockázatának csökkentése, hogy egy AI-munkafolyamat nem megfelelő felhasználóhoz, projekthez, workspace-hez vagy artifacthoz tartozó információt használjon fel.

5. RAG- és forrásbiztonság

Az Axion forrásalapú és retrieval-augmented munkafolyamataiban a visszakeresett információkat scope-, forrás- és hozzáférési korlátozások mellett kezeli.

A rendszer ellenőrizheti a lekérdezéshez használható artifactokat és workspace-forrásokat, azok felhasználói és projektkontextusát, aktív és elérhető állapotát, valamint a lekérdezéshez ténylegesen rendelkezésre álló bizonyítékokat és kontextust.

Az evidence-alapú munkafolyamatok célja többek között annak csökkentése, hogy hiányzó vagy nem hozzáférhető források ténylegesen rendelkezésre álló bizonyítékként kerüljenek kezelésre.

6. Feltöltött fájlok biztonsága

Az Axion a felhasználók által feltöltött fájlokat a feldolgozás előtt több biztonsági ellenőrzésnek vetheti alá.

Az alkalmazott kontrollok közé tartozhat a rosszindulatú fájlok és malware észlelése, fájltípus-, MIME-típus- és fájlkiterjesztés-ellenőrzés, az engedélyezett fájlformátumok korlátozása, fájlméret-korlátozás és további feldolgozás előtti validáció.

A többrétegű file-upload kontrollok célja a rosszindulatú, nem támogatott vagy váratlan tartalom rendszerbe kerülésével kapcsolatos kockázatok csökkentése.

7. Webalkalmazás- és böngészőbiztonság

Az Axion HTTP- és böngészőszintű biztonsági kontrollokat alkalmazhat a webalkalmazást érő támadások kockázatának csökkentésére.

Ezek közé tartozhat a Content Security Policy (CSP), Cross-Origin Resource Sharing (CORS) korlátozások, trusted-host ellenőrzések, framing és embedding korlátozások, külső erőforrások betöltésének korlátozása, aktív webes tartalmak és objektumok használatának korlátozása, valamint további HTTP security header alapú védelmek.

A policy-k célja, hogy a Szolgáltatás működéséhez szükséges erőforrásokat engedélyezzék, miközben korlátozzák a szükségtelen külső végrehajtási és kommunikációs lehetőségeket.

8. Input-, context-, execution- és output-validáció

Az Axion különböző munkafolyamataiban validációs és policy-alapú ellenőrzéseket alkalmazhat a felhasználói bemenet, az AI számára biztosított kontextus, a végrehajtási műveletek és az előállított eredmények között.

Ezek a kontrollok csökkentik annak kockázatát, hogy manipulált input, nem engedélyezett kontextus vagy nem megfelelő AI-output közvetlenül további privilegizált műveleteket váltson ki.

9. Defence in Depth

Az Axion biztonsági modellje nem egyetlen védelmi mechanizmusra támaszkodik. A Szolgáltatás több egymást kiegészítő kontrollt alkalmazhat a felhasználói bemenet, az AI-munkafolyamatok, a forráshozzáférés, a fájlkezelés, az alkalmazáskommunikáció és a végrehajtási környezet különböző pontjain.

Ennek célja, hogy egyetlen biztonsági kontroll megkerülése vagy hibája önmagában ne eredményezze automatikusan a rendszer további biztonsági határainak megkerülését.

10. Folyamatos fejlesztés és a biztonsági intézkedések korlátai

Az Axion a kockázatok jellegéhez és a Szolgáltatás működéséhez igazodó technikai és szervezési intézkedéseket alkalmaz és fejleszt. A kontrollok a rendszer, a képességek és a fenyegetési környezet változásával együtt módosulhatnak.

Egyetlen informatikai vagy mesterségesintelligencia-rendszer sem tekinthető minden körülmények között teljes mértékben biztonságosnak. Az Axion biztonsági intézkedései ezért a kockázatok megelőzésére, felismerésére, korlátozására és mérséklésére irányulnak, és nem jelentenek abszolút garanciát valamennyi biztonsági esemény megelőzésére.

11. A dokumentum hatálya

Jelen Security Overview magas szintű, nyilvánosan megosztható összefoglaló az Axion biztonsági megközelítéséről. Nem minősül teljes technikai specifikációnak, penetrációs tesztjelentésnek, auditjelentésnek vagy a GDPR 28. cikke szerinti Data Processing Agreement Technical and Organisational Measures (TOMs) mellékletének.

Az Axion biztonsági okból nem tesz közzé olyan részletes konfigurációs, infrastruktúra- vagy védelmi információkat, amelyek indokolatlanul növelhetnék a Szolgáltatás támadási felületét.

Kapcsolat biztonsági vagy adatvédelmi kérdésekben: support@axionaiapp.com

Utolsó frissítés: [DÁTUM]
Verzió: [VERZIÓ]