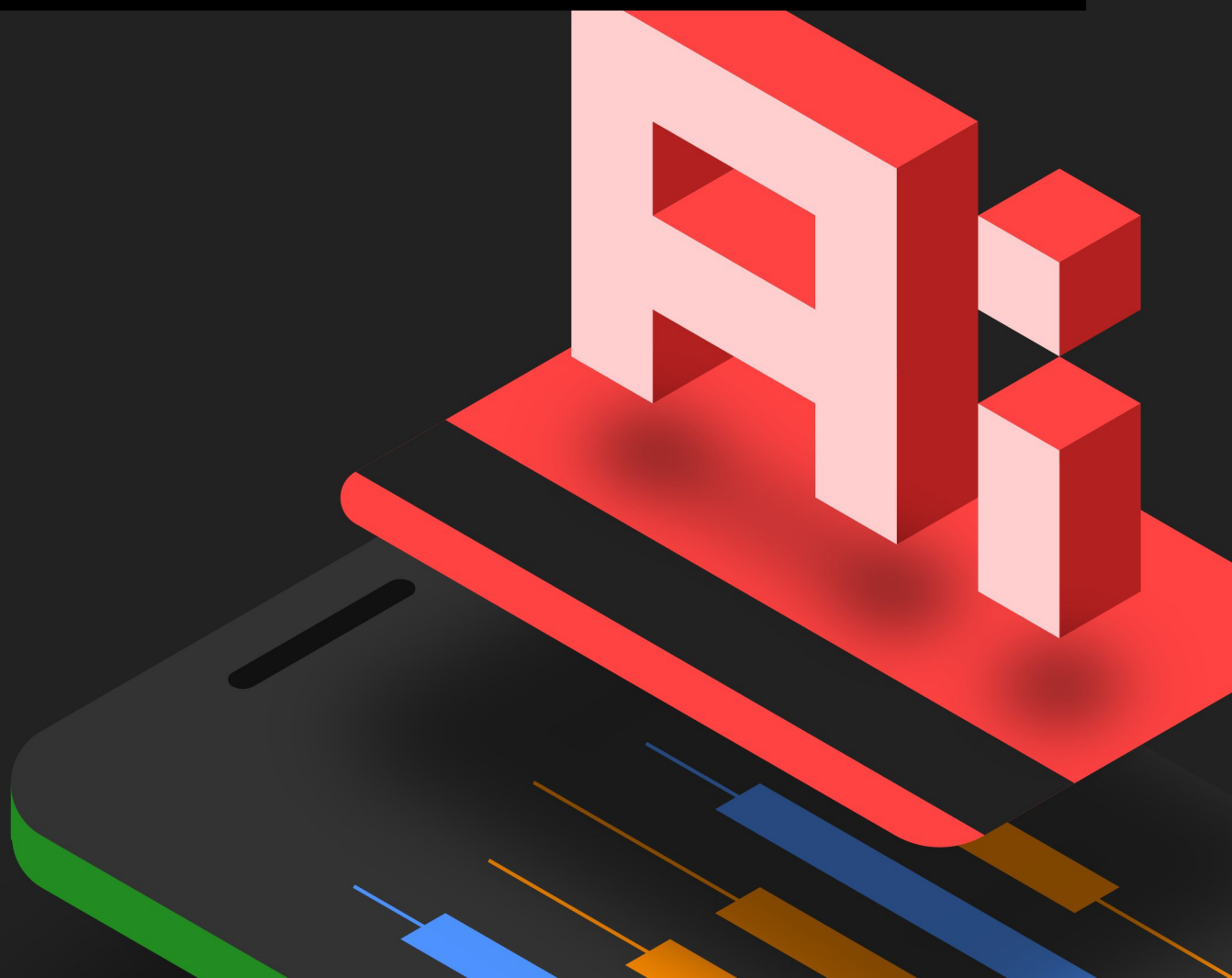
MAKAY

CYBERSECURITY SERVICES & PRODUCTS

AZ AI MEGOLDÁSOK BIZTONSÁGI KOCKÁZATAI A BANKI RENDSZEREKBEN



Bevezetés

A mesterséges intelligencia és a gépi tanulás ma már nem kísérleti technológia a pénzügyi szektorban. A bankok fraud prevention, AML, hitelminősítési, ügyfélszolgálati, ügyfélazonosítási, tranzakcióelemzési és belső döntéstámogató rendszerekben alkalmazzák az AI-alapú megoldásokat.

Ezek a rendszerek jelentős üzleti előnyt adhatnak: gyorsabb döntéshozatal, pontosabb anomáliafelismerés, hatékonyabb ügyfélkiszolgálás és alacsonyabb működési költség érhető el velük. Ugyanakkor az AI-rendszerek más típusú támadási felületet nyitnak, mint a hagyományos alkalmazások. A kockázat nem kizárólag a forráskódban vagy az infrastruktúrában jelenik meg, hanem az adatokban, a modell viselkedésében, az integrációs pontokban, az API-kban és az emberi felügyelet hiányosságaiban is.

A banki AI-rendszerek biztonságát ezért nem elegendő klasszikus sérülékenységvizsgálattal, automatizált scannerekkel vagy általános IT-audittal ellenőrizni. Szükség van AI-specifikus fenyegetési modellezésre, adversarial tesztelésre, LLM-biztonsági validációra, modellkörnyezetek és API-k penetrációs tesztelésére, valamint a DORA és az AI Act elvárásaihoz igazított kontrollértékelésre.

Hol jelenik meg az AI a banki működésben?

A modern pénzintézetekben az AI nem egyetlen elkülönült rendszerként jelenik meg, hanem több üzleti és technológiai folyamatba beépülve.

Terület	Példa AI-funkció	Tipikus biztonsági kérdés
Faud prevention and detection	Valós idejű tranzakcióelemzés, anomáliadetektálás, viselkedéselemzés	Manipulálható-e a modell úgy, hogy csaló tranzakciókat legitimnek minősítsen?
AML / pénzmosás elleni elemzés	Gyanús mintázatok, kapcsolati hálók, strukturálatlan adatok elemzése	Megkerülhetők-e a detektálási szabályok vagy a modell döntési határai?
Hitelminősítés és kockázati scoring	Ügyfélkockázat, fizetőképesség, portfóliókockázat becslése	Átlátható, auditálható és támadásokkal szemben robusztus-e a döntési folyamat?
KYC / ügyfélazonosítás	Dokumentumellenőrzés, biometrikus vagy viselkedésalapú validáció	Megtéveszthető-e a rendszer manipulált dokumentummal vagy szintetikus identitással?
Ügyfélszolgálati chatbotok és LLM-ek	Automatikus válaszadás, ügyfélkérések előfeldolgozása	Kinyerhetők-e ügyféladatok vagy belső információk prompt injection útján?
Belső tudásbázis-asszisztensek	Szabályzatok, eljárásrendek, incidensanyagok keresése	Hozzáférhet-e a modell olyan adatokhoz, amelyhez a felhasználónak nincs jogosultsága?
Kiberbiztonsági monitoring	SIEM/SOC riasztások priorizálása, incidenskorreláció	Kiiktatható-e vagy megtéveszthető-e az AI-alapú detektálás?
Compliance és belső kontroll	Dokumentumelemzés, szabályozási gap analysis	Ellenőrizhető-e, hogy a rendszer nem ad téves vagy nem bizonyítható megfelelőségi következtetést?

Új támadási felület: amikor a modell válik célponttá

Az AI-rendszerek támadása nem mindig klasszikus exploit, jogosultságkiterjesztés vagy hálózati behatolás. Sok esetben a támadó a modell bemenetét, tanítóadatait, döntési logikáját vagy integrációs környezetét manipulálja.

Kiemelt kockázatok

1. Data poisoning – tanítóadat-mérgezés

A támadó a tanító-, validációs vagy visszacsatolási adatfolyamba torzított adatokat juttathat. Banki példában ez fraud vagy AML mintázatok fokozatos eltolását jelentheti, amely hosszabb távon csökkenti a detektálás pontosságát.

Üzleti hatás: csaló tranzakciók átengedése, hibás kockázati besorolás, compliance-kitettség.

2. Adversarial evasion – modellmegkerülés

A támadó kis mértékben módosított bemenetekkel próbálja elérni, hogy a modell téves döntést hozzon. Ez megjelenhet manipulált számlaképekben, módosított tranzakciós mintázatokban, dokumentumellenőrzésben vagy viselkedésalapú csalásdetektálásban.

Üzleti hatás: hamis negatív döntések, gyenge fraud/AML detektálás, reputációs kár.

3. Prompt injection és indirect prompt injection

LLM-alapú banki asszisztenseknél a támadó nem feltétlenül közvetlenül utasítja a modellt. Elég lehet, ha egy e-mailben, dokumentumban,

ügyfélüzenetben vagy tranzakciós megjegyzésben rejt el olyan instrukciót, amelyet a modell feldolgoz.

Üzleti hatás: jogosulatlan adatkiadás, belső szabályok megkerülése, téves ügyfélkommunikáció, hibás művelet-előkészítés.

4. Model inversion és membership inference

Célzott lekérdezésekkel bizonyos esetekben következtetni lehet arra, hogy egy adat szerepelt-e a tanítási halmazban, vagy érzékeny mintázatok rekonstruálhatók a modell válaszaiból.

Üzleti hatás: banktitok, üzleti titok vagy személyes adat sérülése.

5. Model/API abuse és jogosultsági hibák

Az AI-modell gyakran API-kon, adatcsatornákon, pluginokon, RAG-rendszereken, adattárházakon vagy core banking integrációkon keresztül működik. A gyenge jogosultságkezelés itt kritikusabb lehet, mint maga a modellhiba.

Üzleti hatás: jogosulatlan adatlekérdezés, túl széles belső hozzáférés, laterális mozgás, auditálhatatlan döntéslánc.

Mit kell ellenőrizni egy banki AI-rendszerben?

1. AI threat modeling

Minden kritikus AI-rendszerre külön fenyegetési modellt kell készíteni. Ennek ki kell térnie az adatokra, a modellre, az API-kra, a jogosultságokra, a beszállítókra, az üzleti döntési pontokra és a visszacsatolási folyamatokra.

2. Adatbiztonság és adatintegritás

A tanító-, validációs és éles működési adatok forrását, jogosultságait, módosításait és életciklusát ellenőrizni kell. Különösen fontos a training data lineage, a naplózás, a hozzáférési kontroll és az adatszivárgási útvonalak vizsgálata.

3. Modellrobusztusság és adversarial tesztelés

A fraud, AML, KYC és dokumentumelemző modelleket célzottan tesztelni kell manipulált bemenetekkel. Nem elég azt mérni, hogy a modell normál esetben pontos-e; azt is vizsgálni kell, hogyan viselkedik támadási szituációban.

4. LLM guardrails és hozzáférési kontroll

LLM-ek esetén szükséges a prompt injection elleni védelem, a kontextus- és dokumentumhozzáférések szigorú kezelése, a kimeneti szűrés, a tool/plugin-jogosultságok minimalizálása és a műveleti engedélyezési pontok beépítése.

5. Független biztonsági validáció

A banki AI-rendszereket legalább bevezetés előtt, jelentős modellváltozás után és időszakosan is függetlenül ellenőrizni kell. A validációnak ki kell terjednie az alkalmazásra, az API-kra, az infrastruktúrára, az adatfolyamokra, a modellviselkedésre és az üzemeltetési kontrollokra.

Szabályozói elvárás: bizonyítható kontroll, nem csak belső meggyőződés

DORA szempontból:

A DORA a pénzügyi szervezetek IKT-kockázatkezelésére, incidenskezelésére, digitális működési ellenállóképességi tesztelésére és harmadik fél szolgáltatók kezelésére helyez hangsúlyt. Egy banki AI-megoldás akkor is DORA-releváns lehet, ha nem önálló szabályozási objektumként, hanem kritikus IKT-folyamat részeként működik: például fraud monitoring, AML elemzés, ügyfélazonosítás, SOC-támogatás vagy kritikus döntéstámogatás esetén.

TLPT szempontból:

A DORA alatti **threat-led penetration testing** követelmények a jelentősebb pénzügyi szervezeteknél külön figyelmet kapnak. Az EBA 2025-ben közzétett szabályozástechnikai standardjai a TLPT-re kijelölendő szervezetek kritériumait, a tesztelők követelményeit, a módszertant, a scope-ot, az eredményeket és a lezárási/remediációs szakaszokat is részletezik.

AI Act szempontból:

Az EU AI Act kockázatalapú logikát követ. A természetes személyek hitelképességét értékelő vagy hitelpontszámát megállapító AI-rendszerek a magas kockázatú kategóriához kapcsolódhatnak, míg a pénzügyi csalás detektálására használt rendszerek az Annex III adott pontjánál kivételként szerepelnek. Az AI Act alkalmazási ütemezése 2026-ban változó szabályozói környezetben van: az Európai Bizottság tájékoztatása szerint az AI Act fokozatosan lép alkalmazásba, miközben 2025–2026-ban egyszerűsítési és halasztási javaslatok is megjelentek.

Hogyan segít a Makay Kiberbiztonsági Kft.?

A Makay Kiberbiztonsági Kft. banki és pénzügyi környezetben is alkalmazható kiberbiztonsági vizsgálatokkal, penetrációs tesztekkel, AI-specifikus validációval és szabályozási felkészítést támogató technikai ellenőrzésekkel segíti az AI-alapú rendszerek biztonságos bevezetését és üzemeltetését.

Szolgáltatás	Mit vizsgálunk?	Banki érték
AI Security Assessment	AI-rendszer architektúra, adatfolyamok, modellhasználat, integrációk	Átlátható AI-kockázati kép
Fraud / AML adversarial tesztelés	Manipulált bemenetek, megkerülési minták, modellrobusztusság	Detektálási gyengeségek feltárása
LLM és chatbot biztonsági validáció	Prompt injection, indirect prompt injection, RAG-jogosultságok, adatkiadás	Ügyfél- és banktitok védelme
API és integrációs penetrációs teszt	AI-modellt kiszolgáló API-k, backendek, jogosultságkezelés	Kritikus technikai hibák feltárása
AI threat modeling workshop	Üzleti folyamat, adat, modell, infrastruktúra, beszállító	Döntéstámogatás bevezetés előtt
DORA / TLPT támogató vizsgálatok	Kritikus IKT-folyamatok, támadási szcenáriók, remediáció	Szabályozói elvárások támogatása
Forráskód- és konfigurációvizsgálat	AI-integrációk, backend logika, adatkezelés, naplózás	Fejlesztési és üzemeltetési kockázatok csökkentése
Remediációs visszamérés	Javítások ellenőrzése, kontrollok újratesztelése	Auditálható lezárás

Ne hagyja, hogy a legfejlettebb banki védelmi rendszere váljon a legkevésbé ellenőrzött támadási felületté.

Kapcsolat

Makay Kiberbiztonsági Kft.

info@makay.net

+36-30-391-3669

https://makay.net