DaTaOnX – Adathitelesítés végig a digitális értékteremtési és ellátási láncokon Blockchain-technológiával

SOLYMOS GYULA
Blockchain Koalíció
gyula.solymos@gmail.com

Kulcsszavak: Blockchain, Hyperledger, adathitelesítés, quantum-ellenállóság, adatalapú döntés, IoT, supply chain, MI, Ai ACT, eIDAS 2.0, eID

Adatalapú működésre és döntésekre alapozott cégeket és szervezeteket építünk, miközben jelentősen meggyengült a bizalom a digitális adatokban, aminek következtében a cégvezetők alig 20 százaléka mer adatok alapján fontos döntéseket hozni.
A mesterséges intelligencia (MI)-rendszerek elterjedésével egyre fontosabb és megkerülhetetlen kérdés az adathitelesség és manipulációmenteség ellenőrizhetőségének megteremtése cégmérettől függetlenül minden iparágban, mert különben könnyen belső és külső hitelességi válságba kerülhetünk, ami akár a cégünk működését is megrengetheti.
A cikkben bemutatásra kerülő DaTaOnX (Distributed Trust for Oen DATA eXchange) adathitelesítő megoldás egy blockchain-technológiára épített interoperábilis rendszer, amely képes az adatok, dokumentumok, médiafájlok, szoftverkódok változatlanságát azok előállítási vagy tárolási helyétől ellenőrizhetővé tenni. Egyfajta backend megoldásként a meglévő IT-rendszerekben is képes megteremteni az adathitelességet és átláthatóságot a teljes adatellátási lánc mentén, valamint az MI-megoldások tekintetében is.
Az adatokba vetett bizalom megteremtése a digitális térben az egyik legnagyobb kihívás, amivel minden IT- és cégvezetőnek szembe kell néznie iparágtól és szervezetmérettől függetlenül!

1. Bevezetés

A blockchain-technológiáról sajnos még manapság is sok szakembernek a kriptoeszközök, az NFT-hez kötött macskás képek, vagy éppen Ronaldo egymilliárdos perrel csúfos véget ért NFT-projektje jut eszébe, ami a blockchain valódi üzleti problémákat megoldó képességéről, az adathitelesítésről eltereli a figyelmet!

Azok a szakemberek, akik viszont átlátnak ezen a felületen, már nagyban dolgoznak azokon a megoldásokon, amelyek blockchain-technológiára épülve valósítják meg a különféle szervezetek közti adatcserét, adatalapú együttműködést, így munkára fogva a blockchain-technológia azon részét, amire kitalálták, és ami miatt a kriptopiac dollármilliárdjait bízta rá sok millió befektető.

Az EU is bekapcsolódott ebbe a fejlesztésbe az EBSI (European Blockchain Services Infrastructure) révén. A diploma- és képzési mikrotanúsítvány-rendszer, valamint a közjegyzői és az SSID identitáskezelő rendszer mellett számos olyan felhasználási eset fejlődik, ami EU-szinten fogja biztosítani a hitelességet, az átláthatóságot. Ennek a rendszernek megvan a helye a blockchain-ökoszisztémában a határokon átívelő adatalapú együttműködésekben, ugyanakkor az üzleti és közigazgatási megoldásokat is célszerű egy olyan reziliens megoldásra alapozni, ami nem függ egy központi entitástól, vagy céljainak esetleges módosulásaitól.

Ezért szükség van egy olyan, az iparágak közti átjárást és együttműködést biztosító rendszerre, interoperábilis blockchain-megoldásra, ahol az adathitelességet és manipulációmentességet, valamint azok átláthatóságát, felhasználását és követését maguk a folyamatban résztvevő szervezetek, vagy azok által közösen elfogadott/felhatalmazott képviselői biztosítják és tudják kézben tartani. Ennek igényét tovább erősíti, hogy az EBSI-rendszer működésébe „beépített” központosított technológiai kontroll, illetve – ahogy nemrég a BME kutatói publikálták – a „root” adminisztrátori jog által az EU-rendszergazdák elvi lehetősége az adatokhoz való hozzáférésre, ami felvethet némi aggályt, amennyiben az üzleti adatainkat és megoldásainkat erre alapoznánk.

A blockchain-technológia hitelességének alapjai sziklaszilárdak – és a technika mai állása szerint quantum-ellenállóak –, amit mi sem mutat jobban, hogy *az eIDAs 2.0, vagy ahogy újabban emlegetik; eID regulációba jogilag is elismerésre kerül.* A leglényegesebb, hogy a „distributed ledger” azaz a főkönyvi rendszerekben tárolt adatokat a bíróságoknak hitelesnek/változatlannak kell elfogadnia, amiből levezethető és *jogilag megalapozható a blockchain-technológiára épített üzleti megoldások hitelessége is.*

Több európai országhoz hasonlóan a magyar közigazgatás és a Magyar Nemzeti Bank is felismerte a blockchain-technológiában rejlő valódi megoldó képességet. A bankokat és biztosítókat összekötő blockchain-megoldás után a folyamatban lévő EMAP- és a eBlokk-projektek ékes példák erre, de a Blockchain Koalíció által is támogatva egyre több innovatív projekt formálódik az iparban, egészségügyben, energetikában és fintech területen is.

** Jelen dokumentum a Híradástechnika folyóirat HTE Infokom különszámában megjelent cikkhez készült, egyéb irányú felhasználásához a szerző nem járul hozzá.*

Ugyan az EU jogalkotása gőzerővel folyik (pl. AI Act), de mivel a technológia jóval előrébb jár, illetve sok esetben egyre nehezebb lehet azok megszegését felfedezni, ezért *a cégeknek is olyan megoldásokat kell alkalmazniuk a digitalizáció és az adatalapú működésük során, ahol nem a jogszabályok betartása, hanem maga a technológia szavatolja a hitelességet, és ennek az új világnak az alapjainak, az ADAT-nak a hitelességét és a megváltoztathatatlanságát.*

2. Az üzletiadat-manipuláció a mindennapunk része

A DAMA (Data Management Association) felmérése szerint a vezetők 20%-a egyáltalán nem bízik az adatokban. Ennél azonban még súlyosabb tény, hogy 40%-uk csak fenntartásokkal bízik saját szervezetének az adataiban, és talán ennek köszönhetően csak 20%-uk mer adatok alapján fontos üzleti döntéseket hozni a saját cége adatai alapján.

Amíg a vezetők nem mernek megbízni A SAJÁT ADATAIKBAN, addig hogyan várjuk el, hogy megbízzanak a partnereik digitális adatában és az ADATOKRA ÉPÜLŐ MESTERSÉGESINTELLIGENCIA-MEGOLDÁSOKBAN?

2.1. ChatGPT és adatmanipulátor MI-megoldások

Adatalapú döntésekre, együttműködésekre és MI-megoldások használatára alapozott cégeket építünk, miközben egyre kevésbé bízhatunk meg a digitális adatokban, melyeket – paradox módon – az MI-megoldásoknak köszönhetően egyre könnyebb manipulálni…

A mesterségesintelligencia-megoldások robbanásszerű elterjedésével egyre kevésbé bízhatunk meg abban, amit látunk. „Fake" videók, képek, generált személyiségek, csak virtuálisan létező sztárok vesznek minket körül, amelyek közül már nem tudjuk megkülönböztetni, mi az, ami igazi és mi az, ami nem. A ChatGPT elterjedésével és mindenki által történő elérhetőségével már a szövegekben sem bízhatunk meg, hogy tényleg az írta-e, aki aláírta, vagy egy MI…

De mi a helyzet az adatokkal?

Sajnos még kevesen ismerték fel, hogy az adataink is óriási veszélyben vannak, miközben ha a ChatGPT-t magát megkérdezzük, hogy milyen visszaélésekre lehet felhasználni, „Ő" válaszolja azt, hogy „könnyen tudok nagy mennyiségű adatot manipulálni".

Javaslom mindenkinek, hogy maga is próbálja ki, hasonlóan, ahogy mi is megtettük, egy termék minőségének értékeléseit tartalmazó 1000+ soros XLS-táblán. Ezt a ChatGP másodperceken belül képes volt úgy manipulálni, hogy az abban lévő értékelési pontszámok összege és átlaga ne változzon, és az értékelések változása a lehető legkisebb legyen (nehezen lehessen észrevenni), miközben minden olyan sor, amiben egy bizonyos termék megnevezése szerepelt, 100%-os értékelésűre változzon!

Gondoljunk csak bele: mi történne, ha egy belső ember vagy hacker a mi adatainkat manipulálná úgy, hogy az bizonyos külső elvárásoknak feleljen meg, és ezen adatok alapján hoznánk meg az üzleti, vásárlási vagy egyéb döntéseinket?

A manipulált adatokon való döntés azt jelenti, hogy miközben azt gondoljuk, hogy megalapozottan, gondosan és objektíven hoztuk meg a döntésünket, valójában valakinek az érdekei szerint döntöttünk…

2.2. Példák az adatmanipulációra és veszélyeire

Sok történet terjed az interneten arról, hogy hagyományos vagy MI-eszközökkel hogyan manipulálják az adatainkat, és milyen visszaéléseket követnek el vele. Néhány esetet érdemes ezek közül megismernünk, hogy ennek veszélyét kellő mértékben érzékeljük

Excel – a potenciálisan a legveszélyesebb szoftver

A mindennapi üzleti folyamatainkban előszeretettel alkalmazzuk az Excel-t mind döntés-előkészítő mind -támogató megoldást. A FORBES magazin cikke már 2013-ban felhívta a figyelmet a pénzügyi rendszerekben használt XLS-táblák adatainak és arra alapozott számítási metódusainak ellenőrizhetetlen módosításának veszélyeire, illetve az azok által okozott konkrét veszteségekre.

Képmanipuláció – nem csak a szemünket téveszti meg

A Science által közölt beszámoló alapján az elmúlt évtizedekben a legelismertebb Alzheimer-kutatók rendszeresen meghamisították az eredményeiket, amivel nemcsak befektetők dollármillióit, tudományos támogatásokat fecsérelték el, hanem a gyógyszerfejlesztéseket is visszavetette. Nemcsak photoshoppal történt képhamisítás merült fel, volt, aki több kutatásban is felhasználta ugyanazokat a képeket, új kísérleti eredményeknek beállítva őket.

Gondoljuk bele, hogy ma szinte mindenkinek elérhető MI-eszközökkel mennyire egyszerű képeket manipulálni, és nemlétező képeket generálni. Ezt a hírek szerint egy hackercsapat ki is használta, és egy egészségügyi szolgáltató képalkotó diagnosztikai képeibe tett bele, illetve törölt azokból rákmarkereket.

Mivel a cég a több tízezer képe közül nem tudta kiszűrni melyik manipulált, és az újbóli képi diagnosztika nem volt opció, inkább fizettek a zsarolóknak, akik a hírek szerint nem voltak profik, csak jól tudták használni a MI lehetőségeit!

Észrevétlen adatbázis-manipuláció – HMS Defender AIS adatbázis hamisítása

Az AIS rendszer 2021. júniusi nyomonkövetési adatai azt mutatták, hogy a HMS Defender és egy holland fregatt, a HNLMS Evertsen közeledik a krími Szevasztopol kikötőhöz, de nem voltak ott… valójában mindkét hajó nagyjából 300 km-re dokkolt Odesszában.

Mivel az oroszok nem tudták, hogy az AIS adatbázisa manipulált adatokat mutat, ezért ez alapján döntötték el, hogy a HMS Defender-t agresszornak tekintik, és a következő napon a nemzetközi vizeken való közlekedését akadályozták, illetve figyelmeztető lövéseket is leadtak

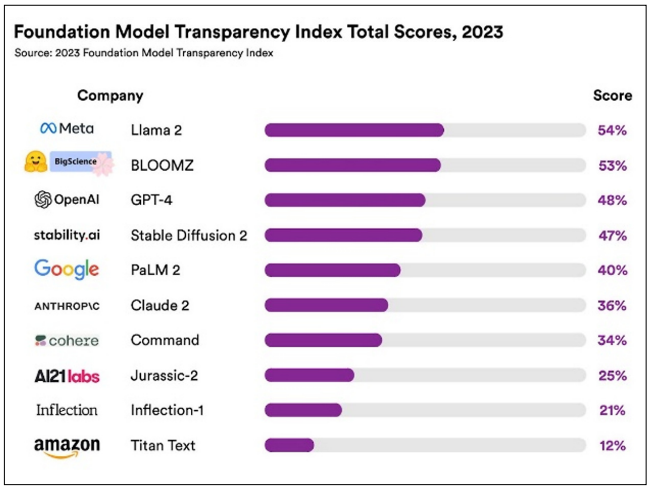
rá, ami a Krimi krízis közepén könnyen annak eszkalációjához vezethetett volna. Az utólagos vizsgálatok szerint az AIS nemzetközileg használt és elfogadott adatbázisát kívülről manipulálták, de arra, hogy ennek ki volt az elkövetője, a mai napig nem derült fény...

Mivel ez az esemény az AIS-rendszer hitelességének megkérdőjelezéséhez vezetett, ezért a hírek szerint annak üzemeltetői olyan adathitelesítési megoldást kezdtek el kidolgozni és abba beépíteni, ami biztosítja, hogy annak adatbázis-manipulációja ne legyen lehetséges.

2.3. MI-modellek átláthatósága

A Stanford Egyetem kutatói jelentést adtak ki a főbb mesterségesintelligencia-modellekről és megállapították, hogy ezekből nagymértékben hiányzik az átláthatóság. Dr. Percy Liang szerint: *„Az elmúlt három évben egyértelmű, hogy az átláthatóság csökkent, miközben a képességek az egekbe szöknek”.*

A TED AI-n aggodalmát fejezte ki az olyan zárt modellek, mint a GPT-3 és GPT-4 felé mutató legújabb trenddel kapcsolatban, amelyek nem adnak kódot vagy súlyokat. Emellett az elszámoltathatósággal, az értékekkel és a forrásanyag megfelelő megnevezésével kapcsolatos kérdéseket is feltett!



A hagyományos és MI-rendszerek működésének átláthatóságának kulcsa is a hiteles adat!

Ezért minden cégnek, szervezetnek, szakigazgatási és állami szereplőnek szükséges is elgondolkodni azon, miként biztosítja a hitelességet az ADAT-alapú üzleti folyamataiban és a mesterségesintelligencia-megoldásokban is!

3. Adathitelesítés szükségessége

Sokan talán még a fenti példák ellenére is szkeptikusak, illetve hamis biztonságérzetük van az adathitelességgel kapcsolatosan, ami azonban a jelen kihívása. Számos IT Security szakértő és trendkutató kongatja a vészharangokat az MI által okozott hitelességi válság tekintetében, közülük két meghatározó hazai szakértő véleményét szeretném ide idézni:

Keleti Artúr:

“...Az ITBN alapítója szerint is a hitelesség a legnagyobb probléma. A mesterséges intelligencia sok mindent művel majd az elkövetkező években, és innentől kezdve amíg létezünk... Ez lesz a jövő, ez nem egy hype ...sokan azt hiszik, hogy elmúlik. Nem fog. A velünk maradó mesterséges intelligencia mindenkinek és mindennek a hitelességét kihívások elé állítja majd és egyben befolyásolja azt is, hogy mit vélünk igaznak!”

(IT Business – ICT-PIAC NAGYKÖNYVE 2023)

Rab Árpád:

“...A digitális tér nagyon fontossá vált, viszont az egymás iránt érzett bizalom csökkent, könnyebb átverni az embereket technológiák segítségével. Könnyű manipulálni..., félrevezetni, ezt nem engedhetjük meg ezért vissza kell szerezni a bizalmat a digitális térben... Ennek legfőbb eszköze a blokklánc-technológia... azt rögzítem, ami van, és ezt nem változtathatja meg senki. Ez kritikus lesz a jövő tekintetében... de az emberek nem mindig fogják tudni, hogy blokklánc-technológiát használnak a háttérben.”

(Karizma Podcast 2022)

3.1. Hol van szükség hiteles adatokra?

A mesterségesintelligencia-megoldások adathitelesség és átláthatóság tekintetében különösen nagy figyelmet érdemelnek, ugyanakkor fontos szem előtt tartanunk, hogy a szervezetünk hiteleségének megtartásához minden olyan adatkörben szükséges megteremtenünk a hitelesség ellenőrizhetőségét, amely(ek)re igaz, hogy:

- kulcsadataink, amelyekre a cégünk üzletmenetét alapozzuk,
- amelyre ellátási láncunkat építjük,
- üzleti döntéseinkhez felhasználunk,
- termékünk minőségét igazolják,
- működésünk mutatóit a piac felé megbízhatóvá teszik,
- amelyeket az adatpiacokon szeretnénk monetizálni.

A fenti felsorolás hitelességmegteremtési fontossági illetve kockázati sorrendet is jelent, amit a cégünk digitalizációja során mindenképpen figyelembe kell vennünk, amikor annak adatalapú működését tervezzük, vagy vesszük górcső alá.

3.2. A ma keletkező adatainkat csak ma lehet hitelesíteni

Az adatok speciális „jószágok”, amelyek hitelessége exponenciálisan csökken, amint elhagyják előállítási helyeiket és eszközeinket (szenzor, IoT-eszköz, adatlap stb.) Ennek oka, hogy az adatkommunikációs, -tárolási és adatcsere-megoldásokba lépve egyre több támadási felülettel érintkeznek, így egyre kevésbé bízhatunk meg azok változatlanságában.

Persze vannak jogi garanciák és technikai megoldások, amelyek a kockázatot csökkenteni hivatottak, de csak akkor lehetünk biztosak abban, hogy az adatainkat nem manipulálták, ha azokat a keletkezésüknél egy olyan technikai megoldással hitelesítjük, amely a teljes életútján végig kiséri.

Az adathitelesség a ma kihívása, a holnap veszélye, de MA keletkező adatokat csak MA lehet hitelesíteni!

3.2.1. Meglévő adatkészletek hitelesítése – adatpiaci megosztás és MI betanítás előtt

Joggal merül fel a kérdés, hogy mit kezdjünk a meglévő adatkészleteink hitelességével?

Ezeket visszamenőleg már nyilván nem tudjuk hitelessé tenni, de erősíteni tudjuk az azokba vetett bizalmat, ha azokat mielőbb hitelesítjük. A meglévő adatkészletek hitelesítésével létrejön egy olyan időpont, amely után minden kétséget kizáróan le tudjuk ellenőrizni, hogy az adott adatkészlet biztosan nem változott meg.

Ennek legtipikusabb alkalmazása az adatbázisaink adatpiaci felajánlásának előkészítése. Ebben az esetben különösen fontos, hogy a vevőm tudja azt, hogy az adott adatkészlet milyen időponttól hitelesített, és ellenőrizhető annak változatlansága.

Amennyiben MI-megoldások betanítására használjuk a meglévő adatkészleteinket, akkor legkésőbb az azzal való feldolgozás előtt szükséges a hitelesítési eljáráson keresztülvinni azokat annak érdekében, hogy az azok segítségével létrejövő MI-modell átláthatóságát és adatkészletének ellenőrizhetőségét biztosítani tudjuk.

3.3. Nyílt adathitelesítési és adatcsere-támogató megoldásra van szükség

Mivel az adatmanipuláció egyre könnyebb és egyre nagyobb veszélyt jelent, ezért – meglátásunk szerint – egy olyan megoldásra van szükség, ami a globális GAIA-X formálódó megoldásának és a különböző iparágakban fejlődő zárt, sokszor nagy belépési küszöbű „siló” blockchain-megoldásoknak (pl. Pharma Ledger) előnyeit egyesíti. Meg kell teremteni annak lehetőségét, hogy létrejöjjön egy olyan hitelesítési megoldás, amihez cégmérettől és adat-előállítási mennyiségtől függetlenül minden szereplő könnyen – akár szolgáltatásalapon – csatlakozni tudjon akár saját hitelesítési csatornát létrehozva.

A DaTaOnX-rendszerünk ennek megvalósítását célozza, amely nem „konkurenciája”, hanem egyfajta interoperábilis fundamentuma lehet az európai üzleti hiteles adatterek és adatcsere-megoldási törekvéseknek, mivel az adat-előállítók tömegeihez és az adat-előállításhoz legközelebb tudná megteremteni a hiteles adat alapjait.

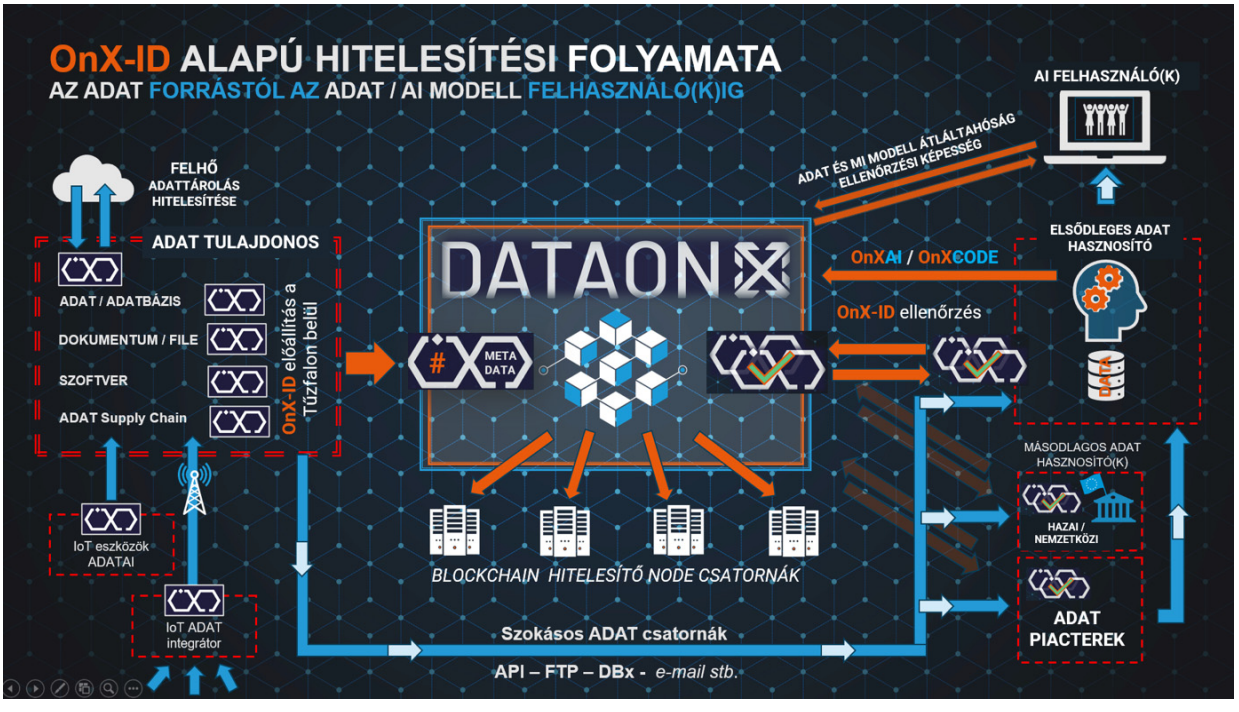
Ezzel mindenki számára elérhetően alacsonyra kerülne az adathitelesség belépési küszöbe, és széles körben létrejöhetne a hiteles adat fogalma, amire az MI-megoldásokat lehetne építeni.

4. DaTaOnX & OnXID alapú interoperábilis adathitelesítés

Az INFOKOM 2023 konferencián bemutatott Distributed Trust for Oen DATA eXchange azaz DaTaOnX-rendszerünk egy Hyperledger Fabric alapú Enterprise blockchain-alapú interoperábilis adathitelesítő megoldás, amelyben az adatok változatlanságának ellenőrizhetőségét az abban tárolt OnX-azonosítók és az adatcserében érintett szereplők – vagy azok megbízottjai – közösen biztosítják.

Mivel a hitelességet az adatnak az ADAT-tulajdonos belső rendszerében előállított egyedi OnxID-azonosítójának blockchain-tárolása biztosítja, és maga a nyers adat alapértelmezetten nem kerül letárolásra a DaTaOnX-rendszerben, ezért a rendszer a GDPR követelményeinek is megfelel.

A következőkben bemutatjuk a DaTaOnX-rendszer főbb építőköveit és működésének alapvető folyamatát, de – mivel az iparjogvédelmi eljárás előtt áll –, ezért annak részleteit jelen cikkben nem publikáljuk.



A leírásban szereplő ADAT fogalma minden olyan egyedi vagy egymással összefüggő adatsorozatot, dokumentumot, média- és egyéb fájlt, szoftverkódot, digitálisan tárolt állományt jelent, amelyből az OnXID előállítható.

4.1. Hitelesítési Backend szolgáltatás/megoldás

A DaTaOnX egy szolgáltatás-alapon igénybevehető, vagy egy ellátási lánc által akár licenszelhető olyan megoldás, amely egyfajta backendként képes beépíteni a hitelességet a meglévő IT-rendszereink és folyamataink mögé.

Nem kell újra írnunk a szoftvereinket, blockchain-megoldásúvá tenni meglévő adatbázisainkat, mert a DaTaOnX háttérrendszerként meghíva ellátja a hitelesítést, és biztosítja, hogy az a teljes ellátási láncunkon végig biztosítani tudja az annak segítségével hitelesített adatainkat. Ezáltal minden partnerünk, a beszállítóink, a vásárlóink a szak- és államigazgatás is meg fog bízni az adatainkban, legyen szó akár termékminőségről, ESG-ről, fenntarthatóságról, vagy akár adatainkra épített finanszírozásól vagy elszámolásokról.

4.2. OnX-azonosító-alapú működés

Az OnX-azonosítók a RAW-adatforrásból, illetve a hozzá kapcsolt eseményekből generált azonosítók, amelyek a DaTaOnX-blockchain adattárában kerülnek tárolásra. Tartalmuk az interoperabilitás biztosítása érdekében előre meghatározott, de több esetben a felhasználási esethez igazítható/testre szabható részeket is tartalmaz, ezen kiegészítések viszont csak akkor lehetnek felhasználhatóak, ha a felhasználó rendelkezik annak definícióival. Ezért ezen felhasználási esetre szabott kiegészítéseket jellemzően egyazon ellátási láncon együttműködő szereplők számára javasolt alkalmazni.

Az OnX ID-generátorok nyílt dokumentációval rendelkező szoftverkódként készülnek el, melyek szabadon felhasználhatóak és beépíthetőek a felhasználók belső rendszereibe, megoldásaiba. Nem kötelező azonban az előre elkészített kódot használni, a logikát a fejlesztők az általuk használt programozási technikával maguk is kifejleszthetik, beépíthetik szoftver- vagy hardver-megoldásaikba.

Az OnX-generátorok a DaTaOnX-rendszerrel előre definiált interfészeken keresztül kommunikálnak. A kommunikáció csak akkor lesz eredményes, ha az adott OnX-generátor a megfelelő felhasználói és biztonsági azonosítókat is elküldi a központi rendszer felé.

4.3. OnX – azonosítók és funkcióik

OnXID – interoperábilis egyedi adat-azonosító

Az OnXiD egy az adatokra egyedileg jellemző HASH-(lásd alább) értékből és az adat keletkezésére jellemző alábbi alapinformációkból áll:

- Adattulajdonos egyedi azonosítója
- Adat keletkezési időpontja (időbélyeg)
- OnXID-előállítás időpontja
- Adatrögzítő eszköz/szoftver/személy egyedi azonosítója
- Adatrögzítés helye (GPS-koordináta)
- Adatpontosság
- DTR-mutató (lásd később)
- Licenc-besorolás
- Üzleti besorolás
- DaTaOnX-csatornaazonosító

HASH egyediségének garanciája

Az OnXID alapja az adatból képzett HASH (SHA 256) egy olyan algoritmussal előállított 256 bites azonosító, amire definíció szerint az alábbi ütközésállóságok jellemzőek:

- „Nehéz” – a gyakorlatban szinte lehetetlen – két olyan bemenetet találni, ami azonos kimenetet eredményez.
- „Nehéz” – a gyakorlatban szinte lehetetlen – egy HASH-értékhez bemenetet találni.
- Lavinaváltozás: a bemenet egy bitjének megváltozása a HASH értékének jelentős (akár 50%-os) megváltozását eredményezi.

OnXID egyedisége és quantum-ellenállósága

A HASH generálása és visszafejtése – a technika jelen állása szerint – quantum-ellenálló, amiből levezetve az OnXID is egyedi és nem feltörhető. Az OnXID egyediségét a blockchainban tárolt ADAT-ot leíró metaadatok tovább erősítik, illetve lehetővé teszik a diszkrét értékek (pl. mérési eredmény) egyedi HASH-képzését is.

4.3.1. OnX kiegészítő azonosítók és funkciói

A kiegészítő azonosítók, az adatról készített OnXID azonosítóhoz kapcsolódnak, de attól elválasztva függetlenül tárolódnak. Ennek köszönhetően az adathitelesség/változatlanság ellenőrzése során az ellenőrző felhasználó az adatnak csak azon aspektusait kezelheti, ismerheti meg, amelyhez az adattulajdonos számára jogot adott.

OnXMETA – adatleíró adatok

Az adat felhasználását támogató leíró adatok, amelyek a különféle felhasználási eseteket támogatják.

OnXTRACK – adat-nyomkövetés

Az eredeti adat változásait (pl. verzió) és hitelességének ellenőrzéseit naplózó azonosító.

OnXLOG – logfile hitelesítés

Szerver és Cyber Security, valamint egyéb logok hitelesítésére.

OnXCODE – szoftver és MI-modell verzió- és CI/CD-folyamatkövetés

A szoftverek speciális verziókövetését és publikációs folyamatát megvalósító azonosító.

OnXATT – Adatattribútum igazolás

Az adatból levezethető attribútum(ok) igazolása – tényadat-átadás nélkül.

OnXSC – Supply Chain összekötés

Egy „ellátási lánc” kapcsolatait leíró és összefogó Smart Contract.

OnXCLR – elszámolóház

Adatalapú elszámolást támogató, az ahhoz kapcsolódó adatokat összefogó Smart Contract.

OnXTWIN – egy adott egyedi termék tulajdonságainak rögzítése és követése

Virtuális termék-létrehozás és -követés megoldása.

OnXAI – Mesterséges Intelligencia betanító adatainak követése

A MI-megoldások betanításához használt adatkészlet hitelességének ellenőrzését és nyomkövetését biztosítja (lásd később részletesen).

OnXAIF – Federált mesterségesintelligencia-megoldásokat betanító adat követése

Célja, hogy a mesterségesintelligencia-algoritmus készítője által nem kontrollált – tűzfalak mögötti – adatkészleteken történő MI-modell-betanítás adathitelességét ellenőrizze, kövesse, a federáció után is visszaellenőrizhetővé tegye.

4.3.2. OnXCON – Adat & Adat összeköttetés

Ennek az azonosítónak a segítségével két hitelesített ADAT OnXID-azonosítója úgy köthető össze, hogy arra csak az adat tulajdonosa, illetve az általa feljogosított felhasználó legyen képes. Ez a megoldás biztosítani tudja például a GDPR-megfelelés szempontjából szükséges anonimizálást abban az esetben is, amikor a DaTaOnX-rendszerben kulcsadatok is tárolásra kerülnek (lásd később).

Az OnXCON-ban nem csak az összekötő azonosító szerepel, hanem egy log is, ami az összekötést megvalósító felhasználókat rögzíti. Ennek segítségével az adatösszekötés ténye megváltoztathatatlanul tárolásra kerül.

4.3.3. OnXDATA

A DaTaOnX-rendszer alapvetően OnXID-azonosítókat tárol, de ha az adott felhasználási eset adatmegosztást tesz szükségessé, akkor a RAW ADAT az OnXDATA-azonosítóba kerül beírásra, ugyanakkor ez esetben is legenerálásra kerül(nek) az OnXID azonosítók.

4.3.4. OnXSC – Smart Contract hitelesítő

A DaTaOnX rendszerben nem csak az adatazonosítókat, hanem az azok felhasználhatóságát, illetve a felhasználási eseteket leíró „Smart Contract” digitális „szerződéseket” is tárolhatunk, amelynek hitelességét és változatlanságát ez az azonosító biztosítja.

Amennyiben az adattulajdonos létrehoz ilyen szerződéseket, úgy egyedileg szabályozhatja az általa hitelesített adatok felhasználóit, felhasználási aspektusait és akár pénzügyi elszámolásait is.

4.4. A DaTaOnX működésének folyamata

4.4.1. Adathitelesítés módja és helye

Az adathitelesítés/változatlanság-igazolás alapja az OnXID és kiegészítőinek blockchain-tárolása.

Hogy ez minél megbízhatóbban igazolja az ADAT változatlanságát – és hogy azt nem lehetett manipulálni –, azt az ADAT-előállító berendezésen belül, vagy az adattulajdonos IT-rendszerének – tűzfallal lehatárolt – keretein belül, annak keletkezéséhez időben és térben a lehető legközelebb kell előállítani.

A DaTaOnX által biztosított OnXID-generáló szoftverkód – az arra alkalmas – szenzorokba, IoT-eszközökbe, vagy azok adatintegrátor/-összesítő berendezéseibe is beépíthető.

Ha ez nem lehetséges, a hitelesítő OnXID generálása a kommunikációs hálózatba érkezés helyén, annak eszközébe is beépíthető, de lehetőség van a fájlszerverbe az adat első letárolása elé is beépíteni azt a folyamatba.

4.4.1.1. DTR – Adathitelesítési távolságmutató

Minél távolabb és később állítjuk elő, annál nagyobb a lehetősége, hogy az adat módosul, illetve manipulálják. Ezért bevezettük és az OnXID-be beépítésre kerül egy *DTR Data Trust Reach* mutató, ami az adathitelesítési adat keletkezésétől való távolságát hivatott megadni.

Értékei:

1. OnXID előállítása az ADAT-előállító eszközben megtörtént.
2. OnXID előállítása az ADAT első kommunikációs csatornába lépéskor történt.
3. OnXID előállítása az ADAT integrátor eszközben/megoldásban történt.
4. OnXID előállítása az ADAT első letárolása előtt történt.
5. OnXID előállítása külső megoldással az első letároláskor hitelesített (pl. digitális aláírás) adatból.
6. OnXID előállítása külső megoldással 2+ éve hitelesített (pl. digitálisaláírás-) adatból.
7. OnXID előállítása az ADAT esetében utólagosan történt 2+ éve változatlan adatbázis alapján.
8. OnXID előállítása az ADAT esetében utólagosan történt 1+ éve változatlan adatbázis alapján.
9. OnXID előállítása már letárolt ADAT esetében utólagosan történt, de 1+ éve változatlan.
10. Az adatkészlet nem hitelesített.

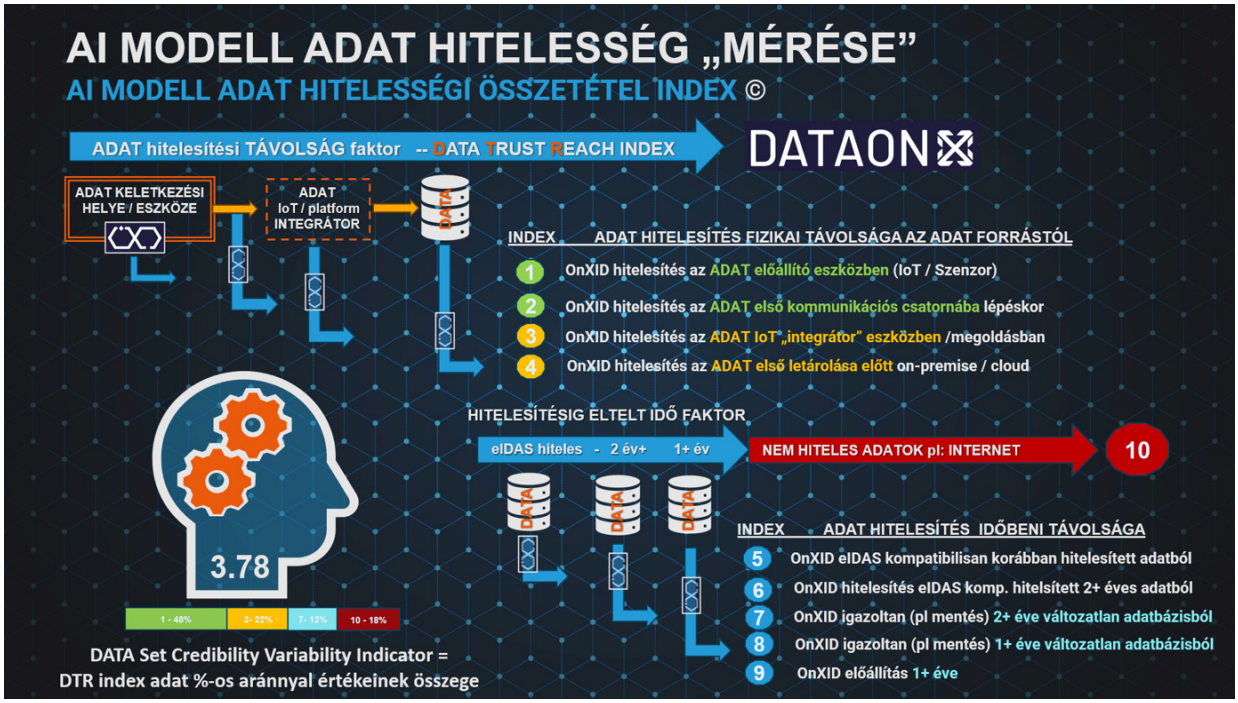
A 7-9. pontok esetében az adatkészlet változatlanságát adatmentéssel, összevetéssel szükséges igazolni.

4.4.1.2. DCVI – ADATKÉSZLET-hitelességi mutató

Mivel a legtöbb estben (pl. MI-megoldásoknál) nem egy-egy konkrét adat hitelességére vagyunk kíváncsiak, ezért bevezettük a *DSCVI–DATA Set Credibility Variability Indicator* mutatót.

Ez súlyozott átlaggal megadja, hogy az adott hitelesített adatkészlet milyen arányban tartalmaz különféle DTR-minősítési adatokat. A DSCVI nem csak megmutatja, hanem össze is láncolja és akár adatról adatra ellenőrizhetővé teszi az adott adatkészlet adatelemeinek hitelességét.

A DCVI mutató lehetőséget ad arra, hogy az MI-modellek ADAT-hitelességhez köthető megbízhatósága mérhetővé és értékelhetővé váljon.



4.4.2. Többcsatornás adathitelesítés

Az OnXID és kiegészítők a DaTaOnX blockchain rendszerében az általa biztosított megváltoztathatatlan módon kerülnek letárolásra. A blockchain-technológia alapja a konszenzusmechanizmusra épül, mely révén az adatokat nem elég összeláncolni, hanem minimum három független *blockchain validátor node* szerveren is le kell tárolni.

Ennek biztosítására a DaTaOnX többcsatornás PERMISSIONED hitelesítő node-szerverrendszerre épít, amelyhez biztosítja, hogy a szereplők több olyan validátorcsoportot alakítsanak ki, vagy olyan meglévő csoporthoz csatlakozzanak, akiket ismernek és akikben megbíznak. Ezek lehetnek magunk az együttműködő cégek, de lehetnek szakigazgatási, vagy állami szereplők is, akikben az adott csatorna adathitelesítési felhasználói közösen megbíznak.

Az adathitelesítési csatornába maga az adattulajdonos vagy felhasználó is beléphet, így maga is biztosítani tudja a hitelességet.

Az egyes csatornák közt lehet átfedés, azaz egy hitelesítő node több csatornának is része lehet.

Azt, hogy egy OnXID melyik blockchain hitelesítési csatornában kerüljön letárolásra, annak előállításakor a csatornaazonosító megadásával kell meghatározni. Egy felhasználónak több csatornába is lehet írási joga attól függően, hogy az adatait milyen területen szeretné felhasználásra megosztani.

4.4.3. Adathitelesség és -változatlanság ellenőrzése

Az adathitelesség/-változatlanság ellenőrzése a felhasználó által a szokásos csatornákon (interfész, adatküldés, mail stb.) eredetiként megkapott adatból történő OnXID előállításán alapul.

Ennek során a felhasználó ugyanazon OnXID-generátort használja, mint amit az eredeti adat hitelesítésekor az adatelőállító, így amennyiben az ADAT változatlan, ennek az ellenőrző OnXID-nak meg kell egyeznie a DaTaOnX-rendszer blockchain tárházában tárolttal.

Az ellenőrző OnXID-t a felhasználó beküldi a DaTaOnX-rendszernek, és amennyiben az megtalálja azt a blockchain adattárházában – és a konszenzusmechanizmus a hitelesítő csatornában is igazolja annak változatlanságát –, visszaigazolja az ADAT változatlanságát.

Adathitelesség kiterjesztett ellenőrzése

Amennyiben az adott felhasználó jogosult az adathoz tartozó további OnX-azonosítókban tárolt adatok lekérdezésére is, úgy a DaTaOnX-rendszer felajánlja azoknak a letöltését is.

Supply Chain-összekötés

Amennyiben az adat egy ellátási lánchoz tartozik, és a felhasználó jogosult azt összekötni, vagy azt összekötni egy általa előállított adattal, úgy ezen operációval folytatódhat a folyamat.

4.4.4. OnXAIT(F) – Mesterséges Intelligencia betanító adatok hitelességének követése

Az OnXAIT egy Smart Contract/chaincode-alapú megoldás, amely a mesterségesintelligencia-modellek betanító adatkészletének OnXID-azonosítóit összeláncolja, és az adott MI-modellhez/szoftverhez rendeli. Az alkalmazott Smart Contract lehetővé teszi a betanító adatbázis növekedésének verzióinak követését is.

Segítségével egy lekérdezéssel ellenőrizhető, hogy a betanításhoz használt adatok mennyiben alapultak hiteles adatokon, milyen DCVI-mutatóval (lásd korábban) rendelkeznek, az adatforráshoz milyen közeli a hitelesítés, így az MI-modellek ADAT-hitelességhez köthető megbízhatósága mérhetővé és értékelhetővé válik.

Az OnXaiT(F) tartalmazza az adatkészlet DSCVI–DATA Set Credibility Variability Indicator mutatót (lásd korább-

ban), amely megadja, hogy az adott MI-modell betanításához használt adathalmaz milyen megbízható, illetve akár adatról-adatra ellenőrizhetővé és átláthatóvá teszi annak adatkészletét.

4.4.4.1. MI-modell verziókövetése

OnXAIT(F) azonosító lánc nem csak az adatkészletet, hanem a modell verzióit is hitelesíti. Az egyes verziókhoz hozzárendeli az azokhoz tartozó adatkészleteket is, így átláthatóvá válik az egyes modellek közti különbség.

MI-modell tesztkészlet-eredmény hitelesítése

Az egyes modellekhez hozzákapcsolható a tesztelésre használt adatkészlet és annak kimeneti eredménye, ami alapján egy adott modell eredményesség szempontjából is átláthatóvá tehető.

4.5. Felhasználási területek

Jelen cikk alapvetően a mesterséges intelligencia adathitelessége köré épül, de alább felsorolunk néhány felhasználási esetet, ahol az adathitelesség fontos lehet, és amely területen a DaTaOnX képességei hasznosíthatóak.

- IoT-eszközben keletkező adatok
- Webes űrlapok – emberi adatbevitel
- Egyedi fájlhitelesítés
- Média/Stream-hitelesítés
- Meglévő adatbázisok hitelesítése
- Felhő-adattárolás ellenőrzése
- Ellátási láncok együttműködésének adathitelesítése, adatmegosztása
- Energiaközösségek elszámolásai
- Kritikus üzemi adatok hitelesítése
- Fintech-adat-alapú hitelértékelés
- Kritikus termékadatok (pl. minőség) hitelesítése
- Drón repülési/permetezési adathitelesítés
- Szoftverkód-hitelesítés
- Szoftver CI/CD-folyamathitelesítés
- Szerver/Security LOG-hitelesítés
- Adat, mint termék hitelesítése
- Adatpiacra kerülő adatok hitelesítése
- Másodlagos adatfelhasználás (pl. hatóság általi) hiteleségbiztosítása

5. Mesterséges intelligencia hiteles adatokon

A szakma nagyon pozitívan fogadta, hogy 2024.03.13-án az Európai Parlament elfogadta az *AI Act regulációt,* aminek része az átlátható MI-rendszerek követelménye is.

A jogi szabályozás ugyan nagyon fontos, mivel az rögzíti a számonkérhetőségi kereteket, de sajnos ez nem elég, mivel nem lehet az egyre nagyobb számban működő összes mesterségesintelligencia-szolgáltatást folyamatosan ellenőrizni!

Sajnos a jogi szabályozás mindig kijátszható és kijátszható lesz, ezért annak megerősítésére olyan technológiát kell használni, ami az AI-megoldásoknál – amelyek az elkövetkező évtizedekben át fogják szőni és meg fogják határozni az üzleti és a társadalmi élet minden területét – ami garantálja az azokban használt adatok hitelességét és átláthatóságát.

A blockchain-technológia és az arra épített DaTaOnX egy olyan informatikai megoldás, amely képes biztosítani, hogy az adatok hitelességét ne a jogszabályok betartása és a rendszergazdába vetett bizalom, hanem maga a technológia garantálja.

6. Epilógus

Napjainkban már minden cég adatokat használ és azokra alapozza döntéseit. Egyre több olyan termékértéklánc és együttműködés épül, ahol meg kell bíznunk egymás adataiban, miközben az adatmanipulációs esetek szaporodása miatt ez egyre nehezebb.

A MI felhasználásával egyre könnyebbé és észrevehetetlenebbé váló adatmanipuláció veszélye egyre nagyobb, így a kulcs- és kritikus adatainknál, valamint a döntéstámogató BI/MI-megoldásoknál elkerülhetetlen az adathitelesítés minden cégméret esetében, mert különben könnyen belső és külső hitelességi válságba kerülhetünk, ami a cégünk működését is megrengetheti!

Az adathitelesség igazolhatósága főként az olyan cégeknél jelent nagy értéket, amelyeknek a terméke maga az adat, vagy az adat a termékhez (pl. minőség, származás stb.) többletértéket ad. Ugyancsak értéket képviselhet, ha az adathitelesítés a cég működésébe (pl. zöldenergia, ESG, fenntarthatóság) vetett bizalmat erősíti.

Egyszerű ADAT-monetizációs törekvés vagy üzleti megfontolás is állhat az adathitelesítés megvalósítása mögött. Azon cégek és szervezetek ugyanis, amelyek az adataikat a keletkezésükkor hitelesítik, értékesek lesznek az MI-megoldásokat fejlesztő cégek számára, így könnyen és magasabb áron fogják tudni értékesíteni az adataikat! Sokan lehet ma még nem tudják, hogy mire lehet az adataikat felhasználni, de elképzelhető, hogy a jövőben egy MI-fejlesztő milliókat fog adni az adatkészletükért, de csak akkor, ha hitelessége minden kétséget kizáróan igazolható!

Sok esetben a kulcsadataink hitelesítése reális alternatíva lehet az IT Security megoldások felskálázásával szemben is. 100%-os Cyber Security-rendszer nem létezik, de a blockchain 100%-ban fogja tudni igazolni, hogy egy-egy hacker-incidens során az adataink kompromittálódnak-e vagy sem!

A fentiek révén ma már minden cégnek és szervezetnek ÉRTÉKET JELENT AZ ADATHITELESSÉG beépítése az adatelőállításba, valamint a belső és Supply Chain együttműködési folyamataiba is.

MI-fejlesztő cégek és az adathitelesség

Az MI-modellek és megoldások hiteles adatra építése és megbízható működése egyre fokozódó piaci elvárás lesz. Ha fejlesztőként ezt elsőként építjük be a megoldásainkba, akkor jelentős piaci előnyre tehetünk szert azokkal szemben, akik kétes vagy nem ellenőrizhető és átlátható adatokra építik a megoldásaikat.

Meglátásunk szerint az MI-rendszerek hiteles működésének 80+ százaléka az adatokban rejlik, amiért a fejlesztő cégek a felelősek, ahogyan az Ai Act előírásainak való megfelelésben is!

Minimális költségek, hosszú távú megoldás

A blockchain-technológia alapú adathitelesítés egy IT-projekt értékének akár 1-2%-ából megvalósítható, a meglévő rendszereinkhez pedig mint egy backend szolgáltatás utólag is könnyen hozzáilleszthető.

Elgondolkodtató, hogy a Bitcoin indulásakor annak rendszerét a banki IT-rendszereknél összemérhetetlenül kisebb költségből indították el – és miközben ma már sok milliárd dollár befektetés biztonságát szavatolja a banki rendszerekkel összemérhető biztonsággal –, az első kibányászott bitcoinok is ugyanúgy hitelesek és elkölthetőek, mint a mai felskálázott rendszerben előálló társaik.

Záró gondolatként mindenkinek érdemes tudatosítani, hogy *a blockchain A HITELESSÉG TECHNOLÓGIÁJA, amely az eID EU reguláció révén jogilag is megalapozottan használható fel az egyre fokozódó adathitelességi elvárások biztosítására minden cég és szervezet esetében.*

A szerzőről

SOLYMOS GYULA az Alpha Management Advisory Kft. tulajdonosa, vezető technológiai szakértője és digitalizációs tanácsadója, valamint a Blockchain koalíció „Ellátási lánc" és Energetikai munkacsoportjának vezetője. Küldetésének tekinti a digitális adathitelesség blockchain-alapú megvalósítását, amelyet elengedhetetlennek tart a megbízható MI-megoldások, az adatalapú cégek, ipari együttműködések és terméklánc, valamint a fenntartható digitális társadalom felépítéséhez és működtetéséhez is. Széleskörű szakmai tapasztalattal rendelkezik az ipari digitalizáció, a blockchain, a mesterséges intelligencia és gépi látás, valamint a 3D/VR- és 5G-technológiák alkalmazásában és az ezekre épülő fejlesztések megvalósításának területén. Tanácsadó cégének szakértőivel olyan Digitális Transzformációs Mesterprogramot/Stratégiát, illetve konkrét IT-megoldásokat dolgoznak ki, amelynek fókuszában nem a technológia, hanem a szervezetek üzleti céljainak elérése áll. Korábban a legnagyobb hazai IT-rendszerintegrátor cég műszaki vezérigazgatójának tanácsadójaként dolgozott. Feladatai mellett több innovatív K+F projektet is kidolgozott, amelyek közül egy kiemelt, több európai tagállam közös érdekét szolgáló EU-program az innovatív egyedi ötlete és az általa kidolgozott know-how révén lett támogatott. 25 éves IT-szakmai életútjának korábbi állomásain több meghatározó hazai cég fejlesztései projektjeit vezette, tanácsadóként támogatta, illetve saját cégének innovatív ötletein dolgozott. Ezek között iparági termékfejlesztési díjjal honorált, valamint a korukat megelőző, ma már mindenki által használt megoldások korai prototípusai is megtalálhatók. Az informatika és digitalizáció területére Pollack Mihály emlékgyűrűvel honorált villamosmérnökként és vállalkozásmenedzser szakmérnökként érkezett. Az élethosszig tartó fejlődés iránti elkötelezettségének köszönhetően jelenleg a negyedik Business Coach diplomáját készül megszerezni.

Hivatkozások

BIANCE NFT's:
https://www.binance.com/en/events/cr7-foreverzone

EBSI:
https://ec.europa.eu/digital-building-blocks/sites/display/EBSI/Home

eIDAS regulation:
https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation

eIDAS 2.0 reguláció javaslat szövege (blockchain-főkönyv):
https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:52021PC0281

Blockchain alapok (Cambridge Business School):
https://www.jbs.cam.ac.uk/faculty-research/centres/alternative-finance/publications/distributed-ledger-technology-systems/

Hyperledger Fabric:
https://www.hyperledger.org/projects/fabric

ESMA Report:
https://www.esma.europa.eu/sites/default/files/2023-10/ESMA12-2121844265-3182_Report_on_the_DLT_Pilot_Regime_-_Study_on_the_extraction_of_transaction_data.pdf

Adatmanipuláció – FORBES: az Excel veszélyei:
https://www.forbes.com/sites/timworstall/2013/02/13/microsofts-excel-might-be-the-most-dangerous-software-on-the-planet/

Képmanipuláció – Alzheimer kutatás:
https://index.hu/techtud/2022/07/27/alzheimer-kor-tudmany-etika-beta-amiloid-csalas-hamisitas-science-ashe-lesne-schrag-selkoe-minnesota-harvard-gyogyszer/

HMS Defender incidens:
https://www.euronews.com/next/2021/06/28/hms-defender-ais-spoofing-is-opening-up-a-new-front-in-the-war-on-reality

Suzuki:
https://maszol.ro/gazdasag/64545-a-suzuki-is-manipulalhatta-a-fogyasztasi-adatokat

Stanford AI-modell átláthatósági indexe:
https://hai.stanford.edu/news/introducing-foundation-model-transparency-index

TED AI Percy Liang előadása:
https://www.ai-event.ted.com/speakers-1/percy-liang-/director-of-the-stanford-center-for-research-on-foundation-models-(crfm)-%26-co-founder%2C-together-ai

AI Act elfogadása:
https://www.europarl.europa.eu/news/hu/press-room/20240308IPR19015/mesterseges-intelligencia-korszakalkoto-jogszabalyt-fogadtak-el-a-kepviselok

CATENA-X autóipari adatellátási lánc:
https://catena-x.net/en/

GAIA-X adatmegosztó hiteles hálózati törekvés:
https://gaia-x.eu/

PharmaLedger – egészségipari hitelesítési rendszer:
https://pharmaledger.org/