



Adatkáosztól az MI-
egyértelműségig: MI-
bevezetés kiváló
minőségű adatokkal



Bevezetés: az MI kiaknázatlan lehetőségei

A mai, gyorsan változó és gyakran kiszámíthatatlan piaci környezetben a generatív és az ügynök-alapú mesterséges intelligencia valódi mérföldkő lehet. Azok a vállalatok, amelyek működésük optimalizálására és összetett kihívások kezelésére törekcsenek, egyre nagyobb mértékben ruháznak be MI-megoldásokba – az ismétlődő feladatok automatizálásától kezdve az ellátási lánc hatékonyságának javításáig. A jelentős potenciál ellenére számos szervezet nem képes a várt üzleti értéket realizálni az MI bevezetéséből. Ennek elsődleges oka gyakran nem az MI-modellek betanításának technikai komplexitása vagy az inferencia költsége, hanem a legértékesebb erőforrás, a vállalati adatok hatékony hasznosításának nehézsége. Mivel az MI-rendszerek alapvetően adatvezéreltek, a hibás vagy nehezen hozzáférhető adatok szükségszerűen pontatlan és kiszámíthatatlan MI-eredményekhez vezetnek. Az MI megbízhatóságának növeléséhez a vállalatoknak modernizálniuk kell adatvagyonukat, biztosítva az adatok skálázható, könnyű hozzáférhetőségét és üzleti kontextusba helyezését.

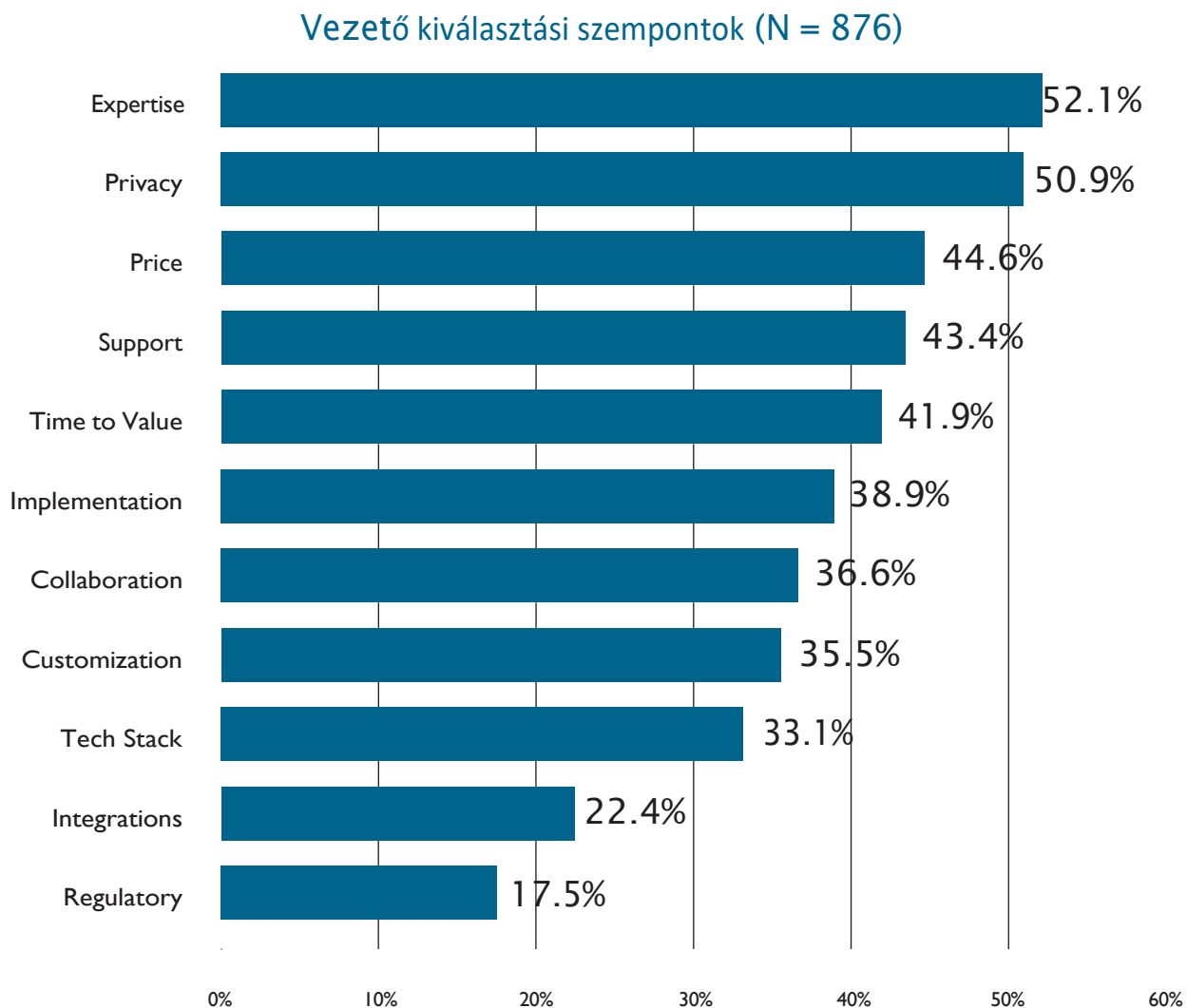
Az MI sikerének leggyakoribb akadályai

Az MI teljes potenciáljának kiaknázásához több kritikus, adatközpontú akadályt kell leküzdni. Jelentős kihívást jelent a rossz adatminőség és a nem megfelelő adatkezelés. A hiányos, inkonzisztens vagy elfogult adatok közvetlen okai a hibás AI-modelleknek, a megbízhatatlan előrejelzéseknek, a rossz döntéshozatalnak, a szabályozási hiányosságoknak, valamint a látható és láthatatlan kockázatoknak való kitettségnek.

A vállalatok hatalmas mennyiségű, potenciálisan hasznosítható adattal rendelkeznek, ugyanakkor ezek mennyisége és összetettsége – különösen strukturálatlan jellegük miatt – komoly kihívást jelent. Az ilyen adatok gyakran dokumentumokban, e-mailekben, prezentációkban és videókban találhatók, ami megnehezíti az MI hatékony alkalmazását. A hagyományos megközelítések, mint a modellek finomhangolása vagy az egyszerű visszakeresésen alapuló, kontextusbővített generálás (RAG), részben segíthetnek, de ritkán képesek az MI teljes értékének kiaknázására. Ráadásul ezek a módszerek gyakran nem illeszkednek jól a strukturált adatok kezelésére szolgáló, bevett adat-hozzáférési megoldásokhoz. Ennek eredményeként a vállalati adatoknak csak kis része kerül időben, következetesen és üzleti hatással a nagy nyelvi modellekbe.

A helyzetet tovább bonyolítja az a tipikus vállalati gyakorlat, amely szerint az idő előrehaladtával széttagolt, egymástól elszigetelt adatkezelő eszközök gyűjteménye alakul ki. Az egyes csapatok és részlegek saját adattárházakat, adat-tavakat, irányítási és integrációs megoldásokat tartanak fenn, ami összetett és nehezen kezelhető adatarchitektúrát eredményez, csökkentve az agilitást és a hatékonyságot. Idővel a technikai adósság és a tudáshiány tovább nő – amint azt a Futurum közel 900 vállalati MI-szakember bevonásával végzett kutatása is alátámasztja: a szakértelem fontosabb szempont az MI-megoldások kiválasztásánál, mint a hagyományos mérőszámok, az adatvédelem vagy a költség (ábra 1).

Ábra 1: MI-beruházások kiválasztási szempontjai



Source: Futurum Research: AI Intelligence Decision Maker: Selection Criteria

Megjegyzés: Amikor az AI-plattformok és -eszközök beszerzési szempontjairól kérdezték, a felmérésben részt vevők több mint fele a szakértelmet jelölte meg elsődleges prioritásként.

Sok szervezet továbbra is szem elől téveszti a lényegét: a generatív MI alkalmazási rétegére – például a chatfelületre és a folyamatlogikára – összpontosít, miközben elhanyagolja az alapvető jelentőségű adat-alapot. Ez megakadályozza, hogy az MI-kezdemenyzések elérjék teljes potenciáljukat, mivel elfedi a háttérben meghúzódó, adatminőséggel kapcsolatos problémákat, mint a pontosság és a relevancia. Amíg a vállalatok nem rendezik ezt az alapvető adat-réteget, a generatív MI-megoldások – beleértve az MI-ügynököket is – nem lesznek képesek beváltani az ígéreteiket.

Új adatparadigma: minőség és kontextus a hatékony MI érdekében

A jó hír az, hogy a strukturálatlan vállalati adatokban rejlő potenciál kihasználásához nem feltétlenül szükséges a meglévő adatkezelési eszközök teljes cseréje. A hangsúlyt érdemes inkább arra helyezni, hogy a modellek minél jobb minőségű, releváns és üzletileg értelmezhető kontextussal rendelkező információhoz jussanak hozzá.

Ez egy átfogó, üzemeltetőközpontú szemlélet (Ops) bevezetését jelenti, amely az adat-fabric architektúrák és gyakorlatok kialakítására összpontosít. A cél az adathozzáférés egységesítése és egyszerűsítése a különböző formátumok és tárolási helyek között. Ez a megközelítés lehetővé teszi a különféle adatfajták zökkenőmentes betöltését, irányítását és visszakeresését – függetlenül attól, hogy azok helyben, felhőben, belső - vagy külső forrásból származnak.

Példaként: az adatszerkezet-architektúrák és -gyakorlatok több kulcsfontosságú elemet egyesítenek.



egységes API-k és hozzáférési módszerek (például modell- és ügynökkommunikációs protokollok),



az SQL mint egységes analitikai interfész emberi és MI-felhasználók számára

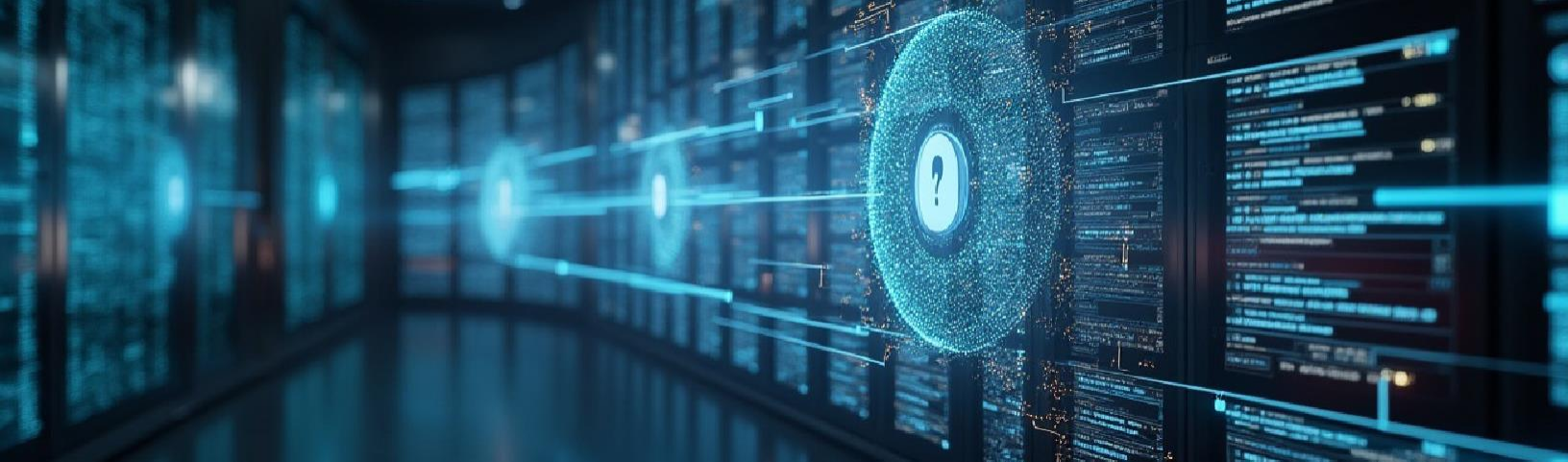


gazdag szemantikus metaadatréteg a strukturált és strukturálatlan adatok feltárásához



nyílt adathozzáférési szabványok és protokollok.

Ezen alapokba történő befektetés hosszú távon növeli az adatminőséget, javítja az integrációs rugalmasságot, csökkenti az üzemeltetési és irányítási költségeket, valamint mérsékli a biztonsági és megfelelőségi kockázatokat.



Megbízható MI az IBM watsonx.data segítségével

Az IBM watsonx.data kifejezetten ezeknek a kritikus követelményeknek a kielégítésére készült, és gyorsan fejlődik az új MI-lehetőségek támogatása érdekében.

Az IBM watsonx.data egy nyílt adat-lakehouse platformként működik, amely adat-fabric képességekkel egészül ki. Az IBM megközelítésének célja az adatok egységesítése, irányítása és üzleti célú aktiválása az eltérő adatbázisok és adatformátumok között. A megoldás egyik legfontosabb megkülönböztető jegye a hibrid architektúra, amely támogatja a helyben telepített környezeteket és a többfelhős működést egyaránt. A kizárólag felhőalapú platformokkal szemben a watsonx.data valódi interoperabilitást biztosít a meglévő adatberuházásokkal, miközben elkerüli a gyártói függőséget. A platform nyílt formátumokban – például Iceberg, Parquet és Avro – tárolt adatokkal dolgozik, amelyek különböző szolgáltatók objektumtárolóiban helyezkedhetnek el.

A 2025 májusában megrendezett IBM Think-en bemutatott, és jelenleg már elérhető watsonx.data jelentős fejlődésen ment keresztül: alapjaiban egyszerűsíti le az MI-t támogató adatarchitektúrát, nagy hangsúlyt helyezve a strukturálatlan adatok kezelésére. Az új képességek lehetővé teszik a strukturálatlan adatok vállalati léptékű betöltését, irányítását és visszakeresését, ezzel megteremtve a pontosabb és skálázhatóbb generatív MI-megoldások alapját.

Az IBM watsonx.data célzott lekérdezőmotorokkal támogatja a munkaterhelések optimalizálását, amelyek közvetlenül a nyílt formátumokban tárolt adatokon futnak. Ilyen például az üzleti intelligenciát kiszolgáló Apache Presto, illetve az adat-engineering feladatokra alkalmas Apache Spark. Ez a többmotoros megközelítés kulcsfontosságú rugalmasságot biztosít, és optimalizálja a teljesítmény–költség arányt a különféle adat-, analitikai és MI-munkaterhelések esetében. Emellett hozzájárul a költséges adattárházi megoldások kiváltásához vagy tehermentesítéséhez, kiegészítve a meglévő környezeteket – például a Snowflake platformot –, különösen az írásintenzív felhasználási esetekben.

A generatív MI támogatására az IBM watsonx.data integrálja a Milvus nyílt forráskódú vektoradatbázist, amelyet a Zilliz fejlesztett, miközben kihasználja az IBM Db2 natív vektortámogatását is. Ezek a képességek tovább bővülnek a nemrég lezárt DataStax felvásárlás technológiáival, amelyek már elérhetők a watsonx.data ügyfelei számára. Az olyan megoldásokkal, mint az AstraDB, a Hyperconverged Database (HCD) és a DataStax Enterprise, az IBM watsonx.data felhasználói a generatív MI szemantikus keresési képességeit ötvözhetik a NoSQL technológiák előnyeivel. Ez lehetővé teszi nagy mennyiségű adat vállalati léptékű kezelését magas rendelkezésre állás és hibatűrés mellett.

A DataStax egyik kulcsfontosságú hozzájárulása a Langflow, egy nyílt forráskódú keretrendszer, amely kifejezetten RAG-alapú adatfolyamokra, MCP-alapú adat- és eszközelérésre, valamint többügynökös alkalmazási munkafolyamatokra lett optimalizálva. A Langflow segítségével a vállalati szakemberek összetett, ügynök-alapú MI-megoldásokat hangolhatnak össze és vezérelhetnek, amelyeket API-végpontként vagy teljes mértékben megfigyelhető, önálló MCP-szerver példányként is képesek éles környezetben elérhetővé tenni.

MI alapú adatkezelés és irányítás

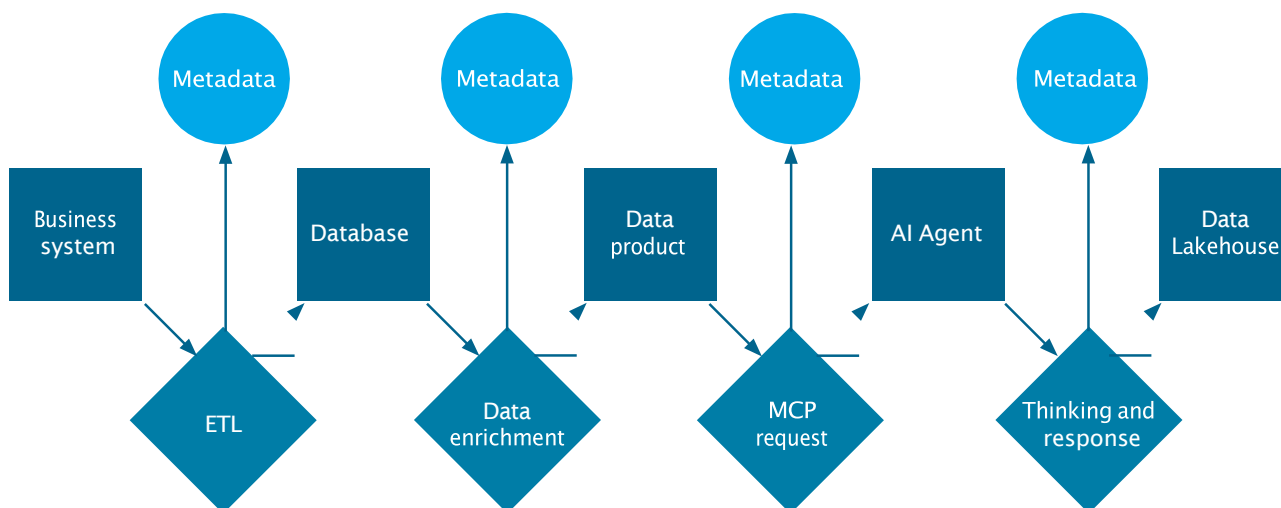
Az IBM watsonx.data kulcsfontosságú megkülönböztető ereje abban rejlik, hogy a platformon belül integrált módon alkalmazza az MI-t, amely érdemben javítja az adatkezelési és adatirányítási (governance) folyamatok hatékonyságát és minőségét.

Ez különösen jól megmutatkozik a IBM watsonx.data új integrációs komponensében, amely azzal az új szemlélettel lép fel, hogy strukturáltságot visz a strukturálatlan adatokba. A puszta vektorizáción túl ez a komponens lehetővé teszi üzletileg értelmezhető, lekérdezhető információk kinyerését. Például képes az entitások automatikus kinyerésére (például az eladási költségek vagy az adásvételi dokumentumok adatainak azonosítására), a személyazonosításra alkalmas adatok (PII) maszkolására, dokumentumok összefoglalására és az adatok osztályozására.

Ezzel a vegyes megközelítéssel az adatszakemberek (és egyre inkább az AI-ügynökök) hatékonyabban tudják alkalmazni a strukturált adatfeldolgozás módszereit: az strukturálatlan adatokból kinyert információkra SQL-lekérdezéseket futtatva pontosabb, determinisztikus számításokat végezhetnek. Ez a megközelítés jelentősen javíthatja az AI-modellek eredményeinek pontosságát, különösen akkor, ha hagyományos RAG-technikákkal kombinálják, amelyek jellemzően nehézkesen használják a strukturálatlan adatokat kontextus biztosítására.

Az integrált adatirányítási képességek — beleértve az adatkatalógust, az adatleszármaztatást (data lineage) és az adatmegosztást — szintén kulcsszerepet játszanak a watsonx.data és a watsonx.data intelligence megoldásokban (2. ábra). Például a generatív és MI-agent megoldások gyakran nem képesek megőrizni és következetesen érvényesíteni a megfelelő hozzáférés-szabályozást a támogató adatfeldolgozási láncokban, különösen akkor, amikor strukturálatlan adatokat használnak forrásrendszerekből, például Boxból vagy SharePointből. Az IBM watsonx.data ezt a problémát úgy kezeli, hogy biztosítja: a watsonx.data felületén keresztül lekérdező felhasználók kizárólag azokat az adatokat érhetik el, amelyekhez az eredeti forrásban meghatározott jogosultságokkal rendelkeznek. Az önmagukban alkalmazott standard RAG-folyamatok és vektoralapú adatbázisok ezzel szemben elvesztik ezt az „őrzési láncot”, leválasztva a vektorizált adatokat az azokhoz kapcsolódó vezérlési mechanizmusokról (fájlrendszer, mappa, dokumentum stb.).

Ábra 2: Adatleszármazás: az adatok útjának feltérképezése a forrástól a célrendszerig



Forrás: Futurum Research

Megjegyzés: Ahogy az adatok a forrástól a megoldásig haladnak, minden egyes állomáson kritikus metaadatok keletkeznek. Amennyiben ezek egy adatkatalógusban láthatóvá válnak, lehetővé teszik annak pontos visszakövetését és rekonstrukcióját, hogy egy agentikus MII-megoldás milyen módon jutott el egy adott válaszhoz.



Összegzés

A sikeres MI-megvalósítások alapját egy stabil, operatív adatplatform képezi, amelyet az adatminőség, a hozzáférhetőség és a rugalmasság jellemez. A nyílt data fabric architektúrák és az ezekhez kapcsolódó működési gyakorlatok folyamatos mérése valós üzleti értéket teremt minden adattípusból — különösen a vállalatoknál felhalmozódott, nagyrészt strukturálatlan adatkészletekből, amelyek túl gyakran maradnak az MI számára elérhetetlenek. A nyílt szabványokra és a rugalmas szoftverbevezetési mintákra helyezett hangsúly lehetővé teszi ezen, eddig kihasználatlan vállalati adatok MI-célú hasznosítását. Emellett a jól szabályozott, biztonságos és együttműködésen alapuló, nagy léptékű adathozzáférés előtérbe helyezése képessé teszi a vállalatokat arra, hogy maximalizálják MI-beruházásaik megtérülését. Ennek eredményeként a szervezetek hatékonyabban tudnak eligazodni a piaci bizonytalanságok között, és új stratégiai lehetőségeket tárhatnak fel.

A tanulmány részletei

SZERZŐK

Brad Shimmin
Vice President & Practice Lead, Data Intelligence,
Analytics, & Infrastructure | The Futurum Group

KIADÓ

Daniel Newman
CEO | The Futurum Group

KAPCSOLAT

Contact us if you would like to discuss this report and
The Futurum Group will respond promptly.

HIVATKOZÁSOK

A tanulmányra az akkreditált sajtóorgánumok és elemzők hivatkozhatnak, azonban kizárólag kontextusban, a szerző nevének, a szerző beosztásának, valamint a „The Futurum Group” megjelölés feltüntetésével. A nem sajtó- és nem elemzői szereplők számára bármilyen hivatkozás kizárólag a The Futurum Group előzetes írásos engedélyével megengedett..

LICENZELÉS

Ez a dokumentum — beleértve minden kapcsolódó háttéranyagot — a The Futurum Group tulajdonát képezi. A kiadvány a The Futurum Group előzetes írásos engedélye nélkül semmilyen formában nem sokszorosítható, nem terjeszthető és nem osztható meg.

JOGI NYILATKOZAT

A The Futurum Group kutatást, elemzést, tanácsadást és konzultációt nyújt számos high-tech vállalat számára, beleértve az ebben a tanulmányban említetteket is. A cég alkalmazottai közül senki sem rendelkezik tulajdonosi részesedéssel a dokumentumban hivatkozott vállalatokban.



IBM

Az IBM hosszú múltra tekint vissza a szervezetek legkritikusabb adatainak és alkalmazásainak kezelésében. Széles körű tapasztalattal rendelkezünk az innovatív vállalati adat- és AI-megoldások területén, beleértve a piacvezető adatbázisokat és a nagyvállalati környezetre felkészített AI-megoldásokat. Ügyfeleinket abban támogatjuk, hogy vállalati adataikat felhőalapú és helyszíni (on-premises) környezetekben egyaránt felszabadítsák és hasznosítsák, és meggyőződésünk, hogy ügyfeleink adatai minden esetben az ügyfeleink tulajdonát képezik.

Futurum

THE FUTURUM GROUP

A The Futurum Group független kutatási, elemzési és tanácsadó vállalat, amely a digitális innovációra, valamint a piacot átalakító technológiákra és trendekre összpontosít. Elemzőink, kutatóink és tanácsadóink nap mint nap világszerte támogatják az üzleti döntéshozókat abban, hogy előre jelezzék iparágaik mélyreható átalakulásait, és a diszruptív innovációt versenyelőnyük megszerzésére vagy megőrzésére használják fel.

Futurum

KAPCSOLAT: The Futurum Group LLC | [futurumgroup.com](https://www.futurumgroup.com) | (833) 722-5337