

AI · TANULMÁNY

Az AI fejlődési irányai és korlátjai a következő években

Gulyás György Tamás • 2026. október 19. • ~45 perc

Koncepcionális tanulmány, 1.0-s változat.

TARTALOM

Absztrakt

Bevezetés: helyzet és kontextus

1 Az adat hitelessége

2 A strukturált kontextus

3 Az auditálható döntéshozatal

4 AI security és governance

Kitekintés

Összegzés

Hivatkozások és további olvasmányok

Absztrakt

Ez a tanulmány koncepcionális, részben spekulatív elemzés arról, hogyan alakíthatja át a mesterséges intelligencia a tudásalapú szakmákat és a szolgáltatóipart a következő öt–tíz évben. Nem a modellekkel és az algoritmusokkal foglalkozom, hanem azokkal a rendszerszintű problémákkal, amelyek az AI tömeges elterjedésével együtt jelennek meg.

Az alapfeltevésem az, hogy a következő évek nehézségei nem az AI képességeiből fakadnak, hanem abból, hogy hiányzik körülötte az infrastruktúra. Minél mélyebben épül be az AI a banki, jogi, egészségügyi, mérnöki és közigazgatási folyamatokba, annál inkább számít, hogy hiteles-e az adat, amiből dolgozik, strukturált-e a kontextus, amit kap, utólag ellenőrizhetők-e a döntései, és kézben tartható-e biztonsági és szabályozási szempontból.

Négy problématerületet vizsgálok. Az első az adat hitelessége egy olyan interneten, ahol a tartalom egyre nagyobb részét gép generálja. A második az AI információhiánya: a rendszerek sokszor nem hibásan következtetnek, hanem hiányos kontextusból dolgoznak. A harmadik az, hogy az AI-asszisztált döntések nem reprodukálhatók, és nem világos a felelősségi láncuk. A negyedik az, hogy a vállalati környezetből hiányoznak az AI-specifikus security- és governance-rétegek.

A négy témát nem külön probléma- és megoldásblokkban tárgyalom, hanem mindegyiket ugyanazon az íven vezetem végig: hol tartunk ma, pontosan mi hiányzik, mi a tét, milyen infrastruktúrával lehetne pótolni, és mit nyerünk vele. A javaslatok: egy digitális proveniencia- és hitelességi réteg, egy strukturált kontextuskezelési modell, egy auditálható döntéstámogató infrastruktúra, valamint egy AI-governance és validációs keretrendszer.

A tanulmány nem szabványjavaslat és nem implementációs útmutató. Azokat a kérdéseket próbálom megfogalmazni, amelyekre a technológiai, üzleti és szabályozási szereplőknek a következő években valószínűleg választ kell adniuk.

BEVEZETÉS

Helyzet és kontextus

A dolgozat célja és műfaja

Ez az írás nem klasszikus tudományos dolgozat, és nem is jövőkutatási vagy hype-anyag. Azt elemzem benne, milyen szerkezeti változásokat indíthat el az AI a tudásalapú szakmákban és a digitális szolgáltatásokban, és ezek milyen problémákat hoznak magukkal.

Azoknak írtam, akik már rutinszerűen használják az AI-t, értik a működését, és a napi munkájuk része. Őket nem az alapok érdeklik, hanem az, hogy merre tart mindez, és milyen hiányzó rendszerekkel kell hamarosan számolniuk.

A kérdést technológiai, gazdasági és infrastruktúra-oldalról is megközelítem. Nem magukat a modelleket vizsgálom, hanem azokat a problémákat és hiányzó rendszereket, amelyek az AI tömeges elterjedésével kezdenek láthatóvá válni.

A központi állításom az, hogy a következő évek nehézségei nem a modellekben jelennek meg, hanem a körülöttük hiányzó infrastruktúrában. Ezért mindegyik témához megoldási irányt is felvázolok.

Az AI ma: általános kép

Az AI-ról szóló vita ma két véglet között mozog. Az egyik szerint a mesterséges intelligencia néhány éven belül átveszi az emberi munka nagy részét, a másik szerint túl van hype-olva, és csak egy újabb eszköz a sok közül. Szerintem egyik sem írja le jól, ami történik, mert a változás máshol zajlik, mint ahol a legtöbben keresik.

Nem a látványos áttörésekben és nem a benchmarkokban látszik, és nem is abban, hogy egy modell hány százalékkal jobb egy teszten. Dokumentumokban, e-mailekben, elemzésekben és ügyintézésben zajlik, ott, ahol a szellemi munka napi része történik.

Egy banki back office-ban, egy ügyvédi irodában, egy kórházi adminisztrációban vagy egy egyetemi tanszéken ma már nem az a kérdés, hogy használnak-e AI-t, hanem az, hogy mennyire épült be észrevétlenül a folyamatokba. Az olyan kérés, hogy „foglald össze”, „ellenőrizd” vagy „javasolj alternatívát”, ma már rutin.

Ezt a változást nem lehet visszacsinálni, és nem csak egyes feladatokat érint: a szakmák és a szektorok szerkezetét is átrendezi.

Az általános célú nyelvi modellek bizonyos feladatokban ma már felveszik a versenyt az átlagos szakemberrel, sokszor jobbak is nála. Strukturált szöveg írása, dokumentumok összefoglalása, fordítás, adatmintázatok felismerése, kódgenerálás és ismétlődő elemzések: ezek ma napi eszközök, nem kísérletek.

Ebből könnyű rossz következtetést levonni. Az AI nem úgy dolgozik, mint egy ember. Nem látja azt a teljes kontextust, amelyben egy szakember mozog: nem ismeri a korábbi döntéseket, a kimondatlan szabályokat, a szervezeti kultúrát és a régi kompromisszumokat. Minden feladatnál úgy viselkedik, mint egy új kolléga, aki szakmailag képzett, de még nem ismeri a környezetét. Ez nem átmeneti hiányosság, hanem a működéséből következik.

Mivel a környezet, amelyben az AI dolgozik, közvetlenül meghatározza az eredményt, a jövőben az lesz a fontos, hogy megbízható, pontos és ellenőrzött adatot kapjon. A válasz minőségét egyre inkább a bemenő adat és az AI számára elérhető információ dönti el, és egyre kevésbé az AI „tudása".

Ebből következik a validáció kérdése is. Az AI adhat olyan választ, amely első ránézésre logikus és szakmailag korrekt, de nem tudja megítélni, hogy az adott üzleti, jogi vagy emberi helyzetben tényleg helyes döntés-e. Egy matematikailag pontos eredmény lehet teljesen érdektelen egy konkrét szervezeti problémában, és egy jól megfogalmazott jogi szöveg lehet használhatatlan egy adott ügyben.

A harmadik tényező a felelősség, és ezt nem lehet megkerülni. Az AI nem vállal felelősséget a döntéseiért, és középtávon nem is fog. Ha egy AI-asszisztált döntés hibásnak bizonyul, a következményt mindig egy ember viseli: az, aki jóváhagyta, aláírta vagy felhasználta az eredményt.

Ez a három tényező (a kontextushiány, a korlátozott validáció és a felelősség hiánya) határozza meg, mi lesz az AI szerepe és a rá épülő iparág helyzete a következő években.

A közeljövő: az AI mint alapfeltétel

A következő egy-két évben az AI-asszisztens a szellemi munka alapeszköze lesz. Ahogy ma az sem versenyképes, aki nem használ e-mailt vagy táblázatkezelőt, rövidesen az AI nélkül dolgozó szakember is hasonló helyzetbe kerül.

Ettől nem lesz mindenki AI-szakértő. Az AI használata a szakmai munka alaprétege lesz, és a különbség nem abban látszik majd, ki használ AI-t és melyiket, hanem abban, ki tudja jól használni. Ki

tudja pontosan megfogalmazni a problémát és átadni az adatokat, ki érti az eredményt, és ki tudja eldönteni, hogy elfogadható-e.

Ezzel együtt elindul egy kevésbé látványos, de mélyebb változás is, amelyről ma már szinte naponta szó esik: átalakulnak a belépő szintű munkakörök. Ha az AI el tudja végezni a klasszikus junior feladatok nagy részét, ezek a szerepek fokozatosan kiürülnek. Rövid távon ez hatékonyabb működést hoz, hosszabb távon viszont szerkezeti problémát: ha nem képződik junior réteg, honnan jönnek a jövő seniorjai?

A senior szakértelme nem egyik napról a másikra alakul ki, hanem évek tapasztalatából, hibáiból és lassú tanulásból. Ha ez az út megszakad, a következménye késve, de biztosan jelentkezik.

Ez nem csak az IT gondja. Ugyanez a minta megjelenhet a jogban, a pénzügyben, az egészségügyben és bármely más tudásalapú szakmában.

A középtávú jövő: a bizalom kérdése

Három-öt éves távlatban a szellemi munka szerkezete tovább változik. A szakembert nem egyetlen AI-rendszer segíti majd, hanem több, egymással együttműködő ágens.

Az egyik adatot gyűjt, a másik elemzi, a harmadik megfogalmazza az eredményt, a negyedik ellenőrzi. A szakember szerepe a közvetlen előállítás felől a koordináció, az ellenőrzés és a végső döntés felé tolódik. Egy digitális csapatot vezet, és a hatékonysága azon múlik, mennyire jól tudja ezt a csapatot megszervezni.

Ennek mélyebb következményei is vannak. Megváltozik a szakmák nyelve. Ma a szellemi munkát végző ember konkrét dokumentumokat ír vagy írat meg az AI-jal, táblázatokat tölt ki vagy töltet ki, folyamatokat szerkeszt kézzel. Néhány éven belül egyre inkább csak a szabályokat és a szándékot írja le, magasabb absztrakciós szinten, és ezt az AI fordítja le konkrét, végrehajtható működéssé.

Egy ügyvéd nem egy szerződést ír meg, hanem egy szabályrendszert, amelyből az AI az adott ügyfélnek elkészíti a konkrét szerződést. Egy orvos nem egy kezelési tervet készít, hanem protokollt ír, amelyet az AI a beteg aktuális állapotához igazít.

Érdemes megnézni, mire használjuk ma az AI-t. Többnyire emberi olvasásra szánt kimenetet készítünk vele: táblázatot, dokumentumot, prezentációt vagy kódot. Ezt elküldjük egy másik embernek, aki egyre gyakrabban szintén AI-jal dolgozza fel. Az egyik AI által generált tartalmat tehát egy másik AI olvassa be, és a kettő közé beékelődik egy embernek formázott köztes lépés, amely egyik gépnek sem a természetes nyelve.

Innen adódik a következő lépés. Eljön az idő, amikor a rendszerek nem emberi formátumban adják át egymásnak az adatot, hanem közvetlenül, gépi reprezentációban, olyan formátumban, amelyet ember már nem feltétlenül tud elolvasni, csak megfelelő AI-eszközzel értelmezni. Ennek első jelei már láthatók a születő ágens-protokollokban: az Anthropic Model Context Protocolja (MCP) és a Google Agent2Agent (A2A) protokollja éppen azt szabványosítja, hogyan adjon át kontextust és feladatot egyik AI-rendszer a másiknak.

Hasonló folyamat indulhat el a programozási nyelvekben is. Ma az AI többnyire a meglévő, embernek tervezett nyelveken dolgozik, de egyre több olyan nyelv és keretrendszer jelenik meg, amelyet kifejezetten nyelvi modellekhez terveztek. Ilyen az LMQL, a Microsoft Guidance és a DSPy, amelyekben a promptolás programozható, strukturált művelet. Könnyen elképzelhető, hogy néhány éven belül lesznek olyan formátumok és nyelvek, amelyeket AI tervez AI-nak, és amelyeknél az emberi olvashatóság már nem elsődleges szempont.

Ez új problémát hoz: születnek hibás döntések, de nem azért, mert az AI rosszul következtet, hanem mert hiányos kontextusból dolgozik.

Az emberi tudás nagy része ugyanis implicit. Nem leírt szabályokból áll, hanem tapasztalatból, háttérismeretből, korábbi kompromisszumokból és kimondatlan feltételezésekből. Az AI ebből csak azt látja, amit kifejezetten megkap. Ha egy kimondatlan szabályt nem kap meg, azt nem is veszi figyelembe.

Ezért egyre többet számít, hogy a kontextust strukturáljuk, és tudatosan beépítsük az emberi ellenőrzési pontokat, mert a rendszer csak olyan jó döntést tud hozni, amilyen teljes a rendelkezésére álló információ.

A dolgozat négy témája

Az AI fejlődéséről szóló beszélgetések ma még főleg a modellek képességeiről szólnak: mennyire pontos egy rendszer, mennyivel jobb az előző generációnál, hány paramétere van, milyen benchmarkokon mit ér el. Ezek jogos kérdések, de egyre kevésbé ezeken múlik a dolog. A technológia eljutott arra a szintre, ahol a legnagyobb nehézségek már nem a modellben jelennek meg, hanem a környezetében.

A kérdés tehát nem az, hogy tud-e az AI jó választ adni, hanem az, hogy milyen adatból dolgozik, mennyire teljes a kontextus, amelyet kap, visszakövethető-e a döntése, manipulálható-e a működése, és ki vállalja a felelősséget, amikor téved.

A következő négy fejezet négy ilyen szerkezeti problémát jár körül. Egyik sem egyszerű technológiai hiba; mindegyik azt jelzi, hogy hiányzik egy infrastruktúra-réteg, amely nélkül az AI hosszú távon nem

lesz megbízható, ellenőrizhető és a társadalom működésébe valóban beépíthető.

A fejezetek ugyanazt az ívet követik: hol tartunk ma (helyzet), mi pontosan a hiányzó réteg (a probléma), mi a tét, milyen infrastruktúrával lehetne pótolni (a javaslat), és mit nyerünk vele (miért működik).

1. TÉMA

01 Az adat hitelessége

Helyzet

A mai AI-rendszerek adatból dolgoznak. Ez nem új felismerés: a gépi tanulás mindig annyira volt jó, amennyire a tanítóadata. A különbség az, hogy korábban az internet tartalmának döntő része emberektől származott, és ez gyorsan változik.

A probléma

A webet egyre nagyobb mennyiségben töltik meg AI által generált szövegek, képek és videók. Egy részük hasznos, egy részük közepes, egy részük egyszerűen zaj. Önmagában az, hogy az AI tartalmat generál, még nem baj. Az a baj, hogy a következő generációs modellek már ebből a tartalomból tanulnak.

Ez önmagát erősítő folyamat. Ha romlik a tanítóadat, romlanak a modellek is, a gyengébb modellek még több gyenge tartalmat állítanak elő, és ez később visszakerül a tanítóadatba. A rendszer így nem stabil egyensúly felé tart, hanem fokozatos minőségromlás felé.

A kérdés tehát az, hogyan különböztetjük meg a hiteles adatot a gépileg generált zajtól.

Mi a tét

Ez elméleti problémának hangzik, de a gyakorlati következményei gyorsan megjelennek. Az orvosi kutatás publikált eredményekre épül, a jogi elemzés precedensekre és dokumentumokra, a pénzügyi modellek piaci adatokra és elemzésekre. Ha ezek között egyre több a gép által generált, ellenőrizetlen vagy torz tartalom, akkor maga az alap válik bizonytalanná.

A digitális világ sokáig abból indult ki, hogy ami olvasható vagy látható, az valahogy a valósághoz kötődik. Ez a feltételezés megszűnőben van.

Ma már nemcsak képeket és videókat lehet hihetően meghamisítani, hanem hangot, dokumentumot, beszélgetést és teljes online identitást is. Hamarosan a „láttam róla videót” és a „hallottam a hangját” típusú állítások semmit sem fognak bizonyítani.

A jövőben valószínűleg nem az az információ lesz értékes, amit könnyű előállítani, hanem az, amelynek a hitelessége bizonyítható. Ezzel együtt megjelenik az igény a validált, visszakövethető, ember által hitelesített adat infrastruktúrájára.

A javaslat: a digitális hitelesség infrastruktúrája

Azt, hogy az AI tud-e valósághű tartalmat előállítani, ma már adottnak tekinthetjük. A kérdés az, hogyan lehet bizonyítani, hogy valami valódi.

Ebben a helyzetben a digitális tartalom önmagában semmit sem bizonyít, és a hangsúly a tartalomról fokozatosan az eredetére kerül át. Egy videónál nem az számít, hogy létezik, hanem az, hogy bizonyítható-e:

- mikor készült,
- milyen eszközzel,
- milyen módosításokon ment keresztül,
- és kihez köthető.

Ehhez másféle infrastruktúra kell, mint amilyen a mai internet.

A mai web elsősorban tárolja és továbbítja a tartalmat. Szerintem a következő évek egyik legfontosabb infrastruktúrája a proveniencia-réteg lesz: olyan rendszerek, amelyek nem magát a tartalmat próbálják „megvédeni", hanem az eredetét és a hitelességét teszik bizonyíthatóvá.

Az első komoly technológiai irány ebben a C2PA-szabvány. A C2PA a digitális tartalomhoz olyan metaadatot kapcsol, amelyből visszakövethető, hogyan jött létre és hogyan módosult. Egy fényképnél például kiderülhet, milyen eszköz készítette, milyen szoftverek nyúltak hozzá, szerkesztették-e utólag, és milyen lépések történtek a publikálás előtt.

Ez önmagában még nem elég, mert nemcsak az a kérdés, hogy hiteles-e a tartalom, hanem az is, hogy hiteles-e a mögötte álló ember. Az AI egyik nagy hatása valószínűleg az lesz, hogy a digitális személyazonosság fokozatosan leválik a valódi emberről. Egyre kevésbé lesz magától értetődő, hogy valóban azzal beszélünk, akinek a másik mondja magát. Hamarosan egy videóhívás, egy hangüzenet vagy egy teljes online meeting sem bizonyít önmagában semmit.

Itt jelenik meg a következő lehetséges infrastruktúra-réteg: a biometrikus és a viselkedésalapú validáció kombinációja.

Egyszerű arcfelismerés vagy hangazonosítás nem elég, ezeket az AI viszonylag gyorsan utánozni fogja. Többrétegű hitelességi modell kell, amely egyszerre vizsgálja a biometrikus mintázatokat, a viselkedési jellemzőket, az interakciós mintákat, az eszköz azonosságát és a teljes digitális eredetláncot.

A digitális eredet és a hitelesség első ránézésre technológiai kérdés, de legalább ennyire gazdasági is.

A mai interneten a tartalom előállítása olcsó, a hitelesség ellenőrzése drága. Az AI ezt a különbséget sokszorosára növeli: a digitális tartalom előállításának költsége a következő években a nullához közelít, a hitelesség bizonyítása pedig egyre nehezebb és egyre értékesebb lesz.

Ebből új piac lesz. Az AI-modellek és a generálóeszközök piaca mellett kialakul egy teljes bizalmi infrastruktúra-réteg, amely az adatok, tartalmak és digitális identitások hitelességét kezeli. Ebben a modellben a validált eredet az érték, nem maga a tartalom.

Az általam elképzelt megoldás első rétege egy validált AI-tréningadat-platform. Az AI-ipar egyik legnagyobb gondja, hogy a modellekhez óriási mennyiségű adat kell, a jó minőségű, ellenőrzött és jogilag tiszta adat viszont ritka. A legtöbb modell ma az internetről gyűjti a tartalmat, ahol az eredet, a minőség és a jogi felhasználhatóság gyakran bizonytalan.

Ez hosszú távon nem tartható fenn. Ahogy a modellek egyre inkább gép által generált tartalomból tanulnak, úgy nő a hiteles, licencelhető és validált adatforrások értéke. A jövő AI-rendszereinél nem az adat mennyisége lesz a szűk keresztmetszet, hanem a megbízhatósága.

Erre lehet válasz egy olyan platform, ahol intézményi adattulajdonosok (egészségügyi szervezetek, mérnöki vállalatok, jogi adatbázisok, kutatóintézetek, egyetemek, bankok) strukturáltan publikálhatják a saját validált adataikat AI-tréningre. A célcsoport nem a tömegpiac, nem bloggerek és általános oktatási tartalmak, hanem nagy értékű szakmai tudásbázisok.

Egy egyetemi kutatóintézet, egy mérnöki vállalat vagy egy egészségügyi szervezet például olyan hitelesített adatcsomagokat publikálhatna, amelyeknek bizonyítható az eredete, ismert a szerzői háttere, végigment egy validációs folyamaton, és tisztázott a jogi felhasználási modellje. Ez nagyon más, mint a mai „scrape-eljünk le mindent az internetről” gyakorlat.

A platform nemcsak intézményektől fogadhatna validált adatot, hanem magánszemélyektől és más adatelőállító cégektől is, és ezek hitelességét a validációs szint határozná meg. A rendszer másik eleme a felhasználókból álló validátori réteg.

A központi ellenőrzés mellett a modell szakmai validátorok közösségére is épül. A validátorok reputációt építenek, ellenőrzik az adatok minőségét, konzisztenciáját és hitelességét, a rendszer pedig folyamatosan súlyozza, mennyire volt megbízható a korábbi munkájuk. Ez bizonyos szempontból a Stack Overflow reputációs modelljére, a sakk Élő-pontszámára épülő bizalmi rendszerek logikájára vagy az IMDb pontozására hasonlít. Fontos cél, hogy ne alakuljanak ki „adatinfluenszerek”: egy adat nem attól lesz hiteles, hogy egy nagy reputációjú validátor jóváhagyta.

A validátori közösség szerepe túlmegy a platformra feltöltött adatok ellenőrzésén. Elvben bármilyen, ma is elérhető internetes dokumentum, kép, videó vagy más digitális tartalom hitelességét vizsgálhatná. Egy ilyen modellben a tartalom technikai változatlanságát kriptográfiai lenyomat és időbélyeg rögzítené, így bizonyítható lenne, hogy egy állomány egy adott időpont óta nem módosult.

Ebből még nem következik, hogy a tartalom valódi is. A valódiság kérdése itt már a validátori hálózathoz és a validátorok felelősségéhez kötődik.

A rendszerben a hitelesség fokozatosan felépülő bizalmi érték, nem igen-nem állapot. A platform folyamatosan súlyozza a korábbi validációk pontosságát, így egy tartalom hitelessége idővel a technikai metaadatokból és a szakmai közösség együttes bizalmából áll össze. Ebből idővel újfajta digitális bizalmi rendszer alakul ki.

Ha egy frissen megjelent tartalmat több validátor is valódinak minősít, gyorsan elérheti azt a bizalmi szintet, amelytől a rendszer már hitelesnek tekinti.

Ha később kiderül, hogy a tartalom mégsem valódi, romlik azoknak a reputációja, akik hitelesnek minősítették, és ez közvetve a korábban általuk validált tartalmak megítélését is érinti.

A validátori fiók egy konkrét személyhez kötődik, akinek az azonosítása történhet biometrikus és viselkedésalapú módszerrel, akár okmányellenőrzéssel is. Így a validátorok reputációja idővel értékes mutatóvá válik (lásd: Black Mirror).

A platform üzleti modellje többoldalú. Az AI-cégek és a nagyvállalati ügyfelek validált adatcsomagokat vásárolhatnak, a bevétel pedig automatikusan megoszlik a platform, az eredeti adattulajdonos és a validátorok között. Ez egyszerre ösztönöz a jó minőségű adatok publikálására és a validációs réteg fenntartására.

A koncepció technikai rétege egy biometrikus KYC-ra épülő, felhős kulcskezelési infrastruktúra. Ebben a modellben a C2PA-aláíró kulcs soha nem hagyja el a felhős HSM-et. A regisztráció banki szintű KYC-folyamattal történik, a hozzáférés pedig biometrikus validációhoz kötött. A biometrikus adatot viszont nem tárolnám központilag: hosszú távon csak zero-knowledge proof alapú megoldás

működhet, amelyben maga a biometrikus adat nem kerül központi adatbázisba. A kulcs így nem egy eszközhöz, hanem egy validált emberi identitáshoz kötődik.

Ez új irányt nyithatna a digitális hitelesség kezelésében.

Miért működik

Hosszú távon nem a tartalom lesz az érték, hanem a bizonyítható emberi eredet. Furcsa módon a „valódiság" lehet az AI-korszak egyik legértékesebb erőforrása.

Ennek gazdasági és társadalmi következményei is lehetnek. Ahogy a digitális tér egyre nagyobb része generálható géppel, úgy jelennek meg azok a környezetek, ahol a validált emberi jelenlét prémium kategória.

Egyes online platformok különbséget tesznek majd hitelesített és nem hitelesített tartalom között, egyes vállalatok és intézmények pedig előírják, hogy bizonyos folyamatokban csak validált eredetű dokumentumot vagy kommunikációt lehet használni.

Ez arra hasonlíthat, ahogyan a HTTPS a web természetes része lett. A korai interneten a hitelesített és titkosított kapcsolat nem volt alapértelmezés, ma nélküle szinte elképzelhetetlen egy szolgáltatás.

Szerintem a digitális eredet és az emberi hitelesség terén is ez fog lejátszódni. A jövő internetén nem az lesz a kérdés, mit látunk online, hanem az, hogy elhisszük-e, hogy valódi. Ahogy a gép által generált tartalom mennyisége exponenciálisan nő, úgy lesznek egyre értékesebbek a licencelt, validált és bizonyítható eredetű adatok rendszerei.

2. TÉMA

02 A strukturált kontextus

Helyzet

Az AI-rendszerek egyik legnagyobb korlátja sokszor nem a logikai képesség, hanem a kontextus hiánya.

Az emberi kommunikáció nagy része implicit. Egy szervezetben rengeteg olyan szabály, háttérismeret és kompromisszum létezik, amelyet senki nem ír le, mégis mindenki figyelembe vesz a döntéseinél, ahogy az etika szabályait és a mindenki által ismert társadalmi normákat is. Az AI ebből csak azt látja, amit megkap.

A mai prompt-alapú működés gyakran abból indul ki, hogy egy probléma minden lényeges kontextusa átadható egyetlen kérdésben vagy utasításban. A legtöbb összetett szakmai döntés ennél jóval rétegzettebb: szabályokra, kivételekre, szervezeti gyakorlatra, előzményekre és emberi mérlegelésre épül egyszerre.

Egy tapasztalt ügyvéd például nem csak a jogszabály alapján dönt. Figyelembe veszi az ügyfél kockázattűrését, a bírósági gyakorlatot, a korábbi ügyeket, és sokszor olyan informális szempontokat is, amelyeket nehéz leírni. Ugyanez igaz az orvosi döntésekre, a mérnöki tervezésre és a vállalati működésre is.

Ennek nagy része ma egyszerűen nem jut el az AI-rendszerekhez. Ahogy az AI egyre több területen segíti a döntéseket, egyre többet számít, hogy strukturáljuk a kontextust, és tudatosan beépítsük a folyamatokba az emberi ellenőrzési pontokat.

A probléma

Az AI bevezetéséről ma gyakran úgy beszélünk, mintha elsősorban technológiai projekt lenne: egy új modell, egy új Copilot, néhány integráció, pár workshop, és a szervezet máris produktívabb.

A valóság inkább arra hasonlít, ahogyan húsz éve az ERP-rendszerek bevezetése zajlott. Papíron ott is minden logikus volt. A rendszer kezelte a beszerzést, a raktárkészletet, a pénzügyet és a gyártást, a demók működtek. A vállalatok mégis újra és újra ugyanabba a problémába ütköztek: nem a technológia bukott el, hanem a szervezet valódi működése ütközött bele a rendszerbe.

Az ERP-bevezetéseknél derült ki igazán, hogy egy vállalat működése sokkal kevésbé strukturált, mint amilyennek kívülről látszik, vagy amilyennek az ott dolgozók addig gondolták. A részlegek ugyanazt a dolgot másképp nevezték, a hivatalos folyamatok és a tényleges működés nem egyezett, a szabályok nagy része nem dokumentumokban élt, hanem emberek fejében és szokásjogban. A kivételek fontosabbak voltak, mint maga a szabványfolyamat. Ez abból is látszik, hogy egy új kolléga betanítása akkor is sokáig tart, ha ismeri az ERP-rendszert és dolgozott már hasonló területen, mert a helyi szokásokat és az íratlan belső szabályokat is meg kell tanulnia.

Az ERP-korszak legfontosabb tanulsága végül szervezeti volt, nem technológiai: a működést csak akkor lehet stabilan automatizálni, ha előtte strukturálják és folyamatokban definiálják.

Az AI ugyanezt a problémát hozza felszínre, csak sokkal élesebben. Az ERP-rendszerek még elviselték a szokásjogon alapuló működést. Az emberek telefonáltak, egyeztettek, fejben vagy Excelben tartották számon a kivételeket, vagy egyszerűen „tudták", hogyan működik valójában a szervezet. Az AI viszont nem viseli el a rejtett kontextust.

Ha a fogalmak nem következetesek, ha a szabályok nincsenek tisztázva, ha ugyanaz a folyamat öt dokumentumban öt különböző formában létezik, akkor az AI nem stabil tudásból dolgozik. Ha sok szabály csak szokásjogként él, az AI töredékekből próbál következtetni, és várhatóan rosszul. Utána jönnek a dolgozók panaszai, hogy „az AI egy egyszerű feladatot sem tud megcsinálni". Persze hogy nem, ha a feladatra ezer mód van, de a helyes egyszer sincs leírva.

Ez más kérdés, mint az előző fejezetben, ahol az adat valódisága volt a tét. A vállalatok adatvagyona ma döntően emberek által előállított, hiteles adatokból áll; itt az a nehéz, hogy ezek következetesek, teljesek legyenek, és a megfelelő kontextusban kerüljenek felhasználásra.

Szerintem itt van az AI körüli egyik legnagyobb félreértés. Amikor egy AI hibázik, hajlamosak vagyunk úgy gondolni rá, mintha egy ember hibázott volna: félreértette a problémát, rosszul gondolkodott, buta vagy nem elég kompetens. Sokszor sokkal egyszerűbb a helyzet: az AI abból dolgozik, amit lát. Olyan, mint egy pályakezdő junior, aki csak abból tud dolgozni, amit elmondtak neki.

Ez hatalmas feladatnak látszik, és sokan lehetetlennek fogják tartani, de nem az. Nem kell sokkal több energiát beletenni, mint egy új kolléga betanításába. Az persze könnyebbnek hangzik, hiszen neki egyszerűen elmondom, mit hogyan szoktunk csinálni. Az AI-nak miért ne lehetne ugyanezt elmondani?

Egy egyszerű példa a kontextushiányra: ha azt mondom egy AI-nak, hogy „írj egy programot, ami összead két számot", néhány másodperc múlva valóban kapok egy működő programot. Jó eséllyel mégsem pontosan azt fogja csinálni, amit szeretnék. Nem azért, mert az AI hibázott, hanem mert a kérésből hiányzik a környezet: milyen rendszerbe készül, milyen hibakezelés kell, milyen platformon

fut, grafikus felületen vagy parancssorban kapja-e az értékeket, oktatási példáról van-e szó vagy egy banki tranzakciós rendszer részéről. A kérés egyszerűen nem tartalmaz elég információt ahhoz, hogy egyetlen helyes megoldás legyen.

Ugyanez történik szinte minden tudásalapú szakmában. Ha valaki azt kéri egy AI-tól, hogy „adj jogi tanácsot a válásomhoz", a rendszer ad tanácsot, csak éppen nem látja az emberi hátteret, az anyagi helyzetet, a gyermekelhelyezési konfliktust, a bírósági gyakorlatot és azokat az informális szempontokat, amelyeket egy tapasztalt ügyvéd ösztönösen figyelembe vesz. Egy orvosi diagnózisnál ugyanígy hiányozhat a teljes klinikai kép, egy vállalati folyamatnál a szervezeti kultúra, egy mérnöki döntésnél a korábbi kompromisszumok története.

Az AI ilyenkor nem rossz választ ad, hanem hiányos világképből próbál következtetni.

Mi a tét

A veszély nem az, hogy az AI önmagában „veszélyes", hanem az, hogy egy bizonyos bonyolultság fölött a hiányzó információ rendszerszintű hibákhoz vezet.

Szerintem a következő években sokan rájönnek majd, hogy az AI fejlődésének következő szakasza nemcsak jobb modellekről szól, hanem arról is, hogy tudunk-e olyan tudás- és kontextus-infrastruktúrát építeni, amelyre ezek a rendszerek megbízhatóan támaszkodhatnak.

Ha a probléma valóban az információhiány, akkor a megoldás sem egyszerűen a „jobb AI". Egy ponton túl nem a modell intelligenciája a szűk keresztmetszet, hanem az, hogy mennyit lát abból a világból, amelyben a döntést meg kell hoznia.

A legtöbb vállalat ma még nem AI-ready. Nem azért, mert nincs AI-ja, hanem mert a működése továbbra is kimondatlan emberi kontextusra épül. A vállalati tudás nagy része Word-fájlokban, PDF-ekben, e-mailekben, meetingjegyzetekben, Excel-táblákban vagy egyszerűen az emberek fejében van. A dolgozók látják a kapcsolatokat, kitalálják a kontextust, és a hiányzó részeket kipótolják a tapasztalatukból. Az AI sokkal rosszabbul boldogul a strukturálatlan tudással.

Ezért a következő években valószínűleg új szervezeti probléma is megjelenik. A technical debt és a data debt mellé odakerülhet a context debt fogalma: az, hogy egy szervezet működésének mekkora része támaszkodik még mindig kimondatlan tudásra, nem dokumentált kivételekre, egymásnak részben ellentmondó folyamatokra, vagy egyszerűen azokra az emberekre, akik „fejből tudják", hogyan működik valójában a rendszer.

Ez több mint dokumentációs probléma, mert közvetlenül eldönti, mennyire tud egy szervezet stabilan AI-t használni.

Elképzelhető, hogy néhány éven belül egy vállalat fontos mutatója a digitális érettség és a cloud readiness mellett egyfajta context health score is lesz. Ez nem a szerverek vagy az infrastruktúra állapotát mérné, hanem azt, hogy a szervezet tudása mennyire következetes és strukturált, mennyi benne a kétértelműség és az ellentmondás, és mennyire értelmezhető AI-rendszerek számára.

A javaslat: a strukturált kontextus infrastruktúrája

Ha a szűk keresztmetszet az információ, és nem a modell, akkor adódik a kérdés: mi lenne, ha nem az AI-t próbálnánk tovább „okosítani", hanem azt a környezetet, amelyből dolgozik?

Mi lenne, ha a szervezeti tudás nem elszigetelt dokumentumokban létezne, hanem strukturált, összekapcsolt és gépileg értelmezhető kontextusként? Ez elsőre technikai problémának tűnik, de inkább szervezeti és kulturális kérdés.

Ebben a szemléletben a hangsúly nem a modellen van, hanem azon a köztes rétegen, amely az emberi szándékot az AI számára feldolgozhatóvá teszi. Új infrastruktúra-réteg jelenik meg a nyers emberi tudás és a végrehajtó AI-rendszerek között; ezt AI-navigátor rétegnek nevezem.

A navigátor réteg sem nem döntéshozó, sem nem végrehajtó rendszer. Köztes környezet, amely strukturálja a kontextust, összekapcsolja az összetartozó fogalmakat, felismeri az ellentmondásokat, és megmutatja, hol hiányos vagy félreérthető az információ.

Ez is AI-réteg, amely folyamatosan figyeli és elemzi az adatvagyont, de nem kérdésekre válaszol és nem döntési javaslatot ad. A dolga az adatvagyon rendszeres karbantartása: az anomáliák, a hiányzó fogalmak és az egymásnak ellentmondó állítások megtalálása és jelzése.

Így a döntéshozó AI nem közvetlenül az embereknek írt, kaotikus dokumentumokból dolgozik, hanem egy előkészített, strukturált tudástérből, amelyet megfelelő megjelenítéssel az ember is használni tud.

Ez azért számít, mert a legtöbb emberi probléma több absztrakciós rétegen át jut el a végrehajtható formáig. Először ott a szándék, az a sokszor pontatlan emberi megfogalmazás arról, mit szeretnénk. Ebből lesz egy strukturáltabb követelményrendszer, aztán egy formális modell, végül a konkrét végrehajtás: programkód, workflow, szerződés vagy automatizált folyamat.

Ez a fordítás mindig is létezett. Egy ügyvéd, egy rendszerszervező vagy egy mérnök folyamatosan absztrakciós szintek között mozog. Az újdonság az, hogy az AI-jal először lehet ezt az átmenetet láthatóvá, strukturálttá és visszakövethetővé tenni.

Amit elképzelek, az tehát nem egy újabb navigátor-chatbot vagy fejlesztői eszköz, hanem élő kontextusrendszer.

A szakember ugyanúgy természetes nyelven dolgozik, mint ma: leírja a problémát, az igényt, a kivételeket és a környezetet. A háttérben viszont több navigátor AI együtt figyeli és strukturálja a folyamatot.

Az egyik navigátor AI valós időben elemzi a szöveget. Felismeri a fogalmakat, észreveszi az ellentmondásokat, jelzi a hiányzó definíciókat, és megpróbálja strukturálni a kontextust. Nem dönt a szakember helyett, inkább olyan, mint egy mindig jelen lévő elemző vagy reviewer.

Egy másik AI már nem az írást segíti, hanem a kész dokumentumból próbál strukturált követelményrendszert készíteni. Egy harmadik konfliktusokat kereshet, egy negyedik compliance- vagy biztonsági problémákat. A rendszer lényege nem egyetlen „szuper-AI", hanem több specializált AI, amelyek ugyanazon a strukturált tudástéren dolgoznak.

A szerkesztett dokumentum és a mögötte lévő tudás közben folyamatosan „okosodik". Nem azért, mert az AI átveszi az irányítást, hanem mert láthatóvá teszi azt a tudást, amely ma rejtve marad a dokumentumok között.

A háttérben valójában nem is hagyományos dokumentumok vannak, hanem közös szemantikus reprezentáció. A szöveg csak nézet; maga a rendszer egy tudásgráf, amelyben fogalmak, szabályok, szereplők és kontextusok kapcsolódnak egymáshoz. Erre azért van szükség, mert az emberi tudás sem lineáris dokumentum. Egy szabály jelentése függhet a környezettől, két állítás részben ellentmondhat egymásnak, és ugyanaz a fogalom mást jelenthet két szervezetben vagy két iparágban. Az ember ezt ösztönösen kezeli, az AI-nak viszont ki kell mondottan ábrázolnia.

Szerintem itt kezdődik a következő generációs vállalati rendszerek infrastruktúrája: olyan rendszerekben, amelyek úgy kötik össze az emberi szándékot a gépi végrehajtással, hogy közben a kontextus nem vész el a rétegek között. Hosszú távon nem az az AI lesz a legértékesebb, amelyik a legjobb választ adja, hanem az, amelyik a legtöbbet érti abból a világból, amelyben a kérdés felmerült.

Miért működik

Az ötlet nem új. Az emberek évezredek óta így kezelik a bonyolultságot: köztes absztrakciós rétegeket hoznak létre.

A jog nem közvetlenül a valóságot írja le, hanem fogalmak, definíciók és szabályok rendszerét építi fel. A mérnöki tervezés sem közvetlenül a fizikai világban történik, hanem modelleken, rajzokon és számításokon keresztül. A szoftverfejlesztésben sem gépi kódban gondolkodunk, hanem specifikációkban, architektúrában, komponensekben és absztrakciókban.

Az összetett rendszerek azért működnek, mert van egy köztes reprezentáció, amelyet az ember és a végrehajtó rendszer is ért.

Az AI-jal először lesz gazdaságilag is megoldható, hogy ezt a réteget folyamatosan karbantartsuk. Korábban egy ilyen tudásrendszer felépítése nagyon drága volt és nehezen tartható fenn. A modelleket, ontológiákat és tudásgráfokat jellemzően kézzel kellett elkészíteni, ezért csak néhány speciális területen terjedtek el, a tartalom folyamatos figyelése és naprakészen tartása pedig gyakorlatilag lehetetlen volt.

Az AI a strukturálás nagy részét már automatikusan el tudja végezni: felismeri a fogalmakat, feltárja a kapcsolatokat, megtalálja az ellentmondásokat, és folyamatosan karbantartja a tudás reprezentációját. Az emberi szakértelmet ezzel nem váltja ki, csak sokkal kevesebb kézi munkával teszi kiaknázhatóvá.

A javaslatom ezért egy olyan infrastruktúra, amelyben az emberi tudás nem dokumentumok, e-mailek és egyéni tapasztalatok töredezett halmazaként létezik, hanem folyamatosan bővülő, közös és gépileg is értelmezhető kontextustérként.

Szerintem nem azok a szervezetek szereznek versenyelőnyt, amelyek elsőként vezetik be az AI-t, hanem azok, amelyek úgy tudják strukturálni a saját tudásukat, hogy az AI valóban megértse.

3. TÉMA

03 Az auditálható döntéshozatal

Helyzet

A modern társadalom nagy része arra épül, hogy a döntéseket utólag ellenőrizni lehet. Egy banki döntést évekkel később is vissza kell tudni követni. Egy orvosi diagnózisnál rekonstruálhatónak kell lennie, milyen információ alapján született. Egy bírósági ügyben elvárás, hogy az ítélet indokolható és reprodukálható legyen.

Az AI mai működése ezzel nehezen fér össze. Egy modell ma ad egy választ; néhány év múlva ugyanaz a modell már nem létezik, a tanítóadat és a környezet megváltozott, és ugyanarra a kérdésre más válasz születik.

A probléma

Hogy ez mekkora gond, az a téttől függ. Egy kreatív szövegnél teljesen elfogadható, hogy az eredmény minden alkalommal más. Egy hitelbírálatnál, biztosítási döntésnél vagy orvosi diagnózisnál viszont súlyos következménye lehet.

A vállalatok ma már szívesen használnának AI-t olyan döntésekhez, mint a hitelbírálat, a jogi mérlegelés vagy az orvosi diagnosztika, de nem merik, mert az eredmény nem reprodukálható, nem auditálható és jogilag nem védhető. Egy 2026-os AI-döntés három év múlva ugyanarra a kérdésre másképp jönne ki, mert a modell is és a kontextus is változott. Bíróság, biztosító vagy szabályozó hatóság előtt ez nem tartható.

A probléma tehát jogi és szervezeti is, nem csak technikai.

Mi a tét

Ma a felelősség végül mindig egy embernél van: annál, aki aláírta vagy jóváhagyta a döntést. A gyakorlatban viszont egyre több helyzetben az ember már csak korlátozottan tudja ténylegesen ellenőrizni az AI által generált eredményt.

Ez veszélyes kombináció: az AI egyre nagyobb szerepet kap a döntésekben, a reprodukálhatóság és az auditálhatóság viszont gyenge marad.

Ahogy az AI mélyebben beépül a szabályozott iparágakba, úgy nő az igény olyan rendszerekre, amelyek rekonstruálni tudják, milyen modell futott, milyen adatból és milyen kontextusban dolgozott, és milyen emberi jóváhagyás történt.

Ha ilyen rendszer nincs, a döntések egyre gyorsabbak lesznek, a mögöttük álló felelősségi lánc viszont szétcsúszik.

A javaslat: az auditálható döntéshozatal infrastruktúrája

A válaszom egy olyan infrastruktúra, amely a döntést végigkövethető folyamatként kezeli, nem egyetlen pillanatként. A kiindulópont maga az ügy.

Egy ügy (egy hitelbírálat, egy bírósági eljárás, egy beteg kezelése) hetekig vagy hónapokig él, mint egy projekt. Egyetlen folyamatos beszélgetésszál, amelybe minden bekerül: dokumentumok, AI-promptok, AI-válaszok, lekérdezések, részdöntések. A lezárás mérföldkövekben történik, előre beállított munkafolyamat szerint, döntési pontokkal.

Ha ugyanazzal az ügyféllel új ügy indul (fellebbezés, reklamáció, ismételt vizit), az előző ügy teljes anyaga csak olvasható bemenetként áll rendelkezésre. Így a kontextus folyamatosan épül, de a régi döntéseket nem lehet visszamenőleg módosítani.

Minden bevitt és lekért tartalom SHA256-tal hash-elve, a Githez hasonló módon kerül mentésre. Ha egy jogszabály vagy belső szabályzat változatlan, csak hivatkozás mutat rá, nem duplikálódik. A kontextus egy hash-lista (manifest), amely a tárolóra mutat. A külső források (jogtár, gyógyszer-adatbázis, BKR-lekérdezés) lekérdezésének időpontja és eredménye is mentésre kerül. A tárhely így kevés, mégis minden visszanézhető.

A reprodukálhatósághoz a használt modellnek évekig pontosan ugyanígy futtathatónak kell maradnia. A publikus felhős modellek (OpenAI, Anthropic API) ezért kiesnek, mert szerződésben sem garantálható a stabilitásuk. A megoldás a befagyasztott (frozen) open-source modell, például Llama, Mistral vagy Qwen: a modellsúlyokat a vállalat eltárolja és birtokolja. A befagyasztott modell soha nem változik, öt év múlva is bitre pontosan ugyanúgy fut.

A reprodukálhatóságnak három szintje van, és fontos eldönteni, melyiket célozzuk.

Eredmény-reprodukálhatóság: ugyanaz a prompt, ugyanaz a kontextus és ugyanaz a modell három év múlva is ugyanazt a választ adja. Ez drága és technikailag nehéz, mert kell hozzá a modellverzió tárolása vagy futtathatósága, a kontextus pillanatképe és determinisztikus mintavételezés.

Folyamat-reprodukálhatóság: látható, hogy akkor és ott mit kérdeztek, mit válaszolt az AI, mit döntött az ember, és milyen adat alapján. Egy mai AI ma más választ adna, de az akkori folyamat rekonstruálható. A célzott piacon (bíróság, ügyvédi iroda, bank) ez önmagában nem elfogadható: a tárgyalóteremben nem hangozhat el, hogy „ma más választ kapnánk".

Indoklás-reprodukálhatóság: a döntés érvelése visszafejthető, vagyis látszik, miért javasolta ezt az AI, és miért fogadta el az ember.

A célzott területeken az eredmény- és az indoklás-reprodukálhatóság együtt a kívánatos. Az első miatt ugyanaz a helyzet évekkel később is pontosan újrajátszható, a második miatt a döntés szakmailag érthető és megvédhető marad.

A folyamat auditálhatósága egyedül nem elég. Attól, hogy pontosan tudjuk, mi történt, még sem az eredmény nem reprodukálható, sem a döntés nem igazolható szakmailag.

Nem minden felhasználás igényel ugyanolyan szintű garanciát. Egy bírósági eljárásnál vagy egy orvosi döntésnél a bizonyíthatóság és a visszajátszhatóság alapkövetelmény, egy üzleti tanácsadási vagy elemzési feladatnál sokkal nagyobb mozgástér is elfogadható. Ezért a rendszer több garanciaszinten is megvalósítható:

1. szint – Teljes reprodukálhatóság, ügyféloldali üzemeltetéssel

A modell és a teljes kontextus a szervezet saját környezetében marad. Célterület: bíróságok, kórházak és gyógyszergyárak, ahol a legmagasabb szintű bizonyíthatóság a követelmény.

2. szint – Platform által garantált reprodukálhatóság

A rendszer menedzselt, több ügyfelet kiszolgáló infrastruktúrán fut, de a modell és a kontextus reprodukálhatósága továbbra is garantált. Elsősorban bankoknak, biztosítóknak és ügyvédi irodáknak felelhet meg.

3. szint – Csak kontextusgarancia

A rendszer nem garantálja, hogy a használt modell változatlan marad, csak azt, hogy a teljes kontextus és az indoklás visszakövethető. Ez elég lehet üzleti tanácsadásra, elemzésre és alacsonyabb kockázatú felhasználásra.

A működési modell mindhárom szinten ugyanarra épül: a rendszer nem elszigetelt kérdés–válasz párokban gondolkodik, hanem egy ügy teljes életciklusát kezeli.

Egy ügyön egyszerre több szakember is dolgozhat. Orvosi esetben konzílium alakulhat, egy jogi ügyön ügyvédi csapat dolgozhat, egy hitelbírálat több jóváhagyási szinten mehet át. A rendszer ezért nem egy felhasználó és egy AI kapcsolatát modellezi, hanem közös, megosztott kontextusteret.

A rendszer bevezetését mindig egy előkészítő (onboarding) szakasz előzi meg. Ebben üzleti elemzők vagy szakterületi szakértők feltérképezik a szervezet döntési folyamatait, és meghatározzák azokat a mérföldköveket, döntési pontokat és jóváhagyási lépéseket, amelyek az adott ügytípusban előfordulhatnak.

Amikor egy konkrét ügy elindul, létrejön a kezdeti kontextusa. Ekkor rögzítik a rendelkezésre álló kiinduló információkat, az ügy résztvevőit, a szerepköröket és az ügy sajátosságait. Ettől kezdve az ügy önálló, folyamatosan bővülő kontextustér.

Az aktív szakaszban kerülnek rögzítésre az új információk, ekkor zajlik a konzultáció az AI-jal, és ekkor születnek a részdöntések. Minden esemény, kérdés, válasz és döntés az ügy kontextusának része lesz, és visszakereshetően megmarad.

A végleges döntéssel az ügy lezárul, és a teljes kontextus befagy. Ettől kezdve a döntést befolyásoló információ már nem módosítható.

Lezárás után is lehet utólag elemezni és magyarázó kérdéseket feltenni. A rendszer válaszolhat arra, miért született egy következtetés, vagy milyen információk vezettek egy döntéshez, de ezek a későbbi kérdések már nem változtatják meg az eredeti kontextust, és nem írhatják felül visszamenőleg a korábbi döntést.

Miért működik

Ez a megközelítés nem azt próbálja elérni, hogy az AI tévedhetetlen legyen, hanem azt, hogy a döntési folyamat átlátható, visszakövethető és reprodukálható maradjon.

A szabályozott szakmákban nem a hibamentesség a követelmény; az emberi döntések sem hibamentesek. Az a követelmény, hogy egy döntést évekkel később is rekonstruálni és szakmailag megvédeni lehessen.

Ha pontosan tudjuk, milyen modell futott, milyen információkból dolgozott, milyen kontextust látott, milyen érveléssel jutott a következtetésre, és ki hagyta jóvá a végső döntést, akkor az AI-asszisztált döntés ugyanabba a felelősségi keretbe illeszthető, mint bármely más szakmai döntés.

Nyitott kérdések és kompromisszumok

Ez a megközelítés nem ad minden problémára azonnali választ, és több kérdést szándékosan nyitva hagyok.

Az egyik a modellek minőségének időbeli változása. A frontier modellek várhatóan gyorsabban fejlődnek, mint a befagyasztott open-source modellek, ezért egy ma megfelelőnek tartott rendszer néhány év múlva gyengének tűnhet az akkori technológiai szinthez képest. Ezt a reprodukálhatóságért fizetjük: a stabilitás és a folyamatos fejlődés között elkerülhetetlen a kompromisszum.

Kérdés az is, hogyan őrizzük meg és futtassuk a modelleket hosszú távon. Ha egy döntést tíz vagy húsz év múlva kell újrajátszani, akkor nemcsak a modell súlyait, hanem a teljes futtatási környezetet is meg kell őrizni. Ez már infrastruktúra- és archiválási probléma is, nem csak szoftveres.

Nyitott kérdés a post-mortem elemzés is. Egy későbbi vizsgálatban szabad-e új kérdéseket feltenni? Ha igen, melyik válasz számít hitelesnek: az eredeti modellé, egy korszerűbb modell új válasza, vagy mindkettő? Ezekre a technológiának, a szabályozásnak és az iparági gyakorlatnak együtt kell választ adnia.

Valószínűleg ezeken a kérdéseken múlik, hogy az AI-ból a következő években bizonyítható, auditálható és jogilag is védhető döntéstámogató eszköz lesz-e.

4. TÉMA

04 AI security és governance

Helyzet

A vállalati AI-rendszerek terjedésével új biztonsági és szabályozási problémák jelentek meg.

A klasszikus IT-security sok területet lefed: hálózatvédelmet, jogosultságkezelést, titkosítást, sérülékenységkezelést. Az AI viszont új támadási felületet hoz: maga a modell lesz támadási felület.

A probléma

A prompt injection például azért veszélyes, mert a rendszer számára az utasítás és az adat gyakran ugyanabban a formában érkezik: természetes nyelven. Egy dokumentumba vagy e-mailbe rejtett utasítás úgy befolyásolhatja az AI működését, hogy a felhasználó észre sem veszi.

Közben egyre több szervezet próbál saját, zárt AI-rendszert építeni. Bankok, kórházak, biztosítók, állami intézmények és gyógyszergyárak nem engedhetik meg, hogy az érzékeny adataik nyilvános modellekbe kerüljenek, ezért gyorsan nő az on-premise AI-telepítések száma.

Ehhez újfajta szakértelem kell. A klasszikus IT-security ismerete nem elég: a szakembernek egyszerre kell értenie a modellek működését, a tanítóadat kockázatait, a prompt-alapú támadásokat, a szabályozási megfelelést és az AI-rendszerek governance-kérdéseit.

Mi a tét

A piac erre ma még nincs felkészülve.

Az AI-security ma széttagolt terület. Vannak szabványkezdemények, governance-keretrendszerek és technikai ajánlások, de nincs egységes, kiforrott infrastruktúra vagy szakmai gyakorlat.

Ahogy az AI egyre érzékenyebb rendszerekbe kerül, várhatóan kialakul egy új iparági réteg, amelynek nem a modellfejlesztés lesz a dolga, hanem az AI-rendszerek auditálása, validációja, biztonsági felügyelete és szabályozási megfeleltetése.

A konkrét nehézség három ponton sűrűsödik: az adatizoláció (az AI ne lásson bele más beteg kartonjába, más ügyfél hitelébe, más bíróság ügyébe), a prompt injection elleni ellenállás és a

szabályozási megfelelés leképezése. Ehhez ma a klasszikus IT-security szakértelem nem elég, és nincs egységes iparági tanúsítvány. A nagy tanácsadók governance-szinten beszélnek a témáról, de a telepítés szintjén hiány van.

A javaslat: AI-governance és validáció

Ebben a piaci résben jelenhet meg egy új, erősen specializált tanácsadói réteg: olyan szolgáltatók, amelyek kifejezetten az AI-rendszerek biztonsági auditjára, validációjára és szabályozási megfelelésére szakosodnak.

A működés kezdetben elsősorban tanácsadás, de ezzel párhuzamosan kialakulhat egy saját, nyilvános és dokumentált szabvány is. Ez meghatározza, milyen követelményeknek kell egy AI-rendszernek megfelelnie az adatizoláció, a prompt injection elleni védelem és a szabályozási megfelelés terén.

Az ügyfél első lépésben részletes gap-elemzést és implementációs javaslatot kap, később független audit keretében megszerezheti a tanúsítványt is. A kettőt szét kell választani: ugyanaz a szervezet nem adhat tanácsot és nem tanúsíthatja a saját munkáját, mert az sértené a függetlenséget és az egész rendszer hitelességét.

A saját szabványt átmeneti eszköznek tekintem. Várható, hogy az iparág a következő években kiforrottabb, hivatalos szabványokat és megfelelőségi keretrendszereket alakít ki: az ISO/IEC 42001 fejlődik, az EU AI Act megfelelőségi módszertana stabilizálódik, és új iparági gyakorlatok születnek. Ezért a saját szabványt eleve úgy érdemes felépíteni, hogy a kontrollpontjai könnyen megfeleltethetők legyenek a későbbi hivatalos követelményeknek. A cél nem az, hogy az ügyfelet bezárjuk egy saját rendszerbe, hanem az, hogy később gond nélkül át tudjon lépni a hivatalos tanúsítási rendszerekbe.

A tanúsítás több elemből állhat. Az ügyfél kap egy tanúsítványt, amelyet beépíthet a compliance-dokumentációjába, valamint egy nyilvánosan ellenőrizhető digitális igazolást és regiszterbejegyzést. Kap emellett részletes kontrollkatalógust és pontozást arról, mely területeken felel meg a követelményeknek, és hol maradt még hiány. Mindezt részletes auditjelentés támasztja alá, amely belső és szabályozói célra is használható.

Egy ilyen tanúsítási rendszer várhatóan több szakaszban épül fel.

Az első szakaszban a hitelesség és a referenciaügyfelek megszerzése a cél. A szabvány és a tanúsítás ekkor még teljesen saját kezdeményezés, ezért az első sikeres projektek és a szakmai ismertség a legfontosabb.

A második szakaszban jöhetnek az együttműködések akkreditált tanúsító szervezetekkel, például a TÜV-vel, a BSI-vel vagy a DEKRA-val. Ilyenkor a specializált szolgáltató adja az AI-specifikus szakértelmet, a partner pedig az akkreditációt és a jogi hátteret.

A harmadik szakaszban egy ilyen szervezet akár önálló, akkreditált szereplő is lehet, és beléphet a magas kockázatú AI-rendszerek hivatalos tanúsítási piacára.

Minden új tanúsítási rendszer ugyanazzal a problémával indul: az első napon senki sem ismeri a pecsétet. Ismertség nélkül nincs piaci értéke, piaci érték nélkül nincsenek ügyfelek, ügyfelek nélkül pedig nem jön létre az ismertség.

Ez nem új jelenség. Az ISO 27001 és a SOC 2 sem egyik napról a másikra lett iparági alapelvárás. Először csak egy megközelítés volt a sok közül, és a piacnak fokozatosan kellett megbíznia benne.

A bizalomhoz több dolog kell. Az egyik a teljes átláthatóság: a szabvány és a kontrollpontok nyilvánosak, ellenőrizhetők és megismételhetők. A másik a név szerint vállalt szakértelem, mert a piac elsősorban az emberekben és a tapasztalatukban bízik. Ugyanígy számítanak az első, nagy presztízsű ügyfelek, és az, hogy a tanúsítás működését időnként független fél is felülvizsgálja.

A nyilvános regiszter is a legitimációt szolgálja. Eleinte csak néhány bejegyzés van benne, később viszont önmagában is bizalmi infrastruktúra lehet.

Egy tanúsítvány akkor lesz iparági tényező, amikor külső szereplők is elismerik: amikor egy szabályozó hatóság elfogadható bizonyítékként hivatkozik rá, amikor egy biztosító kedvezményt ad a tanúsított rendszerekre, amikor az ügyfelek megkövetelik, vagy amikor a versenytársak is hasonló rendszert kezdenek építeni.

Ettől kezdve a tanúsítvány már nem a kibocsátó hitelességéből él, hanem saját jogán bizalmi intézmény.

Miért működik

Ennek a megközelítésnek az az értéke, amit a nagy tanácsadó cégek és a klasszikus IT-security ma jellemzően nem fednek le: a telepítés szintjén ellenőrizhető megfelelés. Az ügyfél nemcsak ajánlásokat és governance-elveket kap, hanem kézzelfogható bizonyítékot arra, hogy az AI-rendszere megfelel bizonyos biztonsági és szabályozási követelményeknek.

Ez ott számít, ahol bizonyítani kell az adatizolációt, a prompt injection elleni védelmet vagy a szabályozási követelmények teljesülését. A tanúsítás így közös nyelv lesz, amelyet az ügyfél, a compliance-terület, a szabályozó és később a hivatalos tanúsító szervezetek is értenek.

További előnye, hogy nem zárja be az ügyfelet egy saját rendszerbe. Ha a saját kontrollrendszer eleve a formálódó iparági szabványokkal való kompatibilitásra épül, akkor később természetes az átmenet a hivatalos keretrendszerek, például az ISO/IEC 42001 vagy az EU AI Act megfelelőségi rendszere felé.

A kockázatokat is ki kell mondani. Nem garantálható, hogy egy saját tanúsítási rendszer valaha széles körű iparági elfogadottságot ér el. Sok hasonló kezdeményezés maradt rétegmegoldás, vagy tűnt el néhány év alatt. Ennek általában két oka van: vagy a piac nem érzi elég súlyosnak a problémát ahhoz, hogy fizessen a megfelelésért, vagy egy nagyobb szereplő gyorsabban rak össze hasonló ajánlatot, és maga köré szervezi a piacot.

A legóvatosabb forgatókönyv szerint a saját szabvány végül a tanácsadás módszertani kerete marad: segíti a közös munkát az ügyfelekkel, strukturálja az auditokat és egységesíti a szakmai megközelítést, de nem lesz önálló piaci szabvány.

Ezt nem tartanám kudarcnak. Már az is ér valamit, ha van egy nyilvános és következetes kontrollrendszer, amely összeköti a mai AI-security gyakorlatot a várható hivatalos megfelelőségi követelményekkel. Ha pedig a piac idővel konszolidálódik, és a módszertan részben vagy egészben beépül a hivatalos tanúsítási rendszerbe, akkor a kezdeményezésből a későbbi iparági szabványok egyik építőköve lehet.

Kitekintés

Hova fektessünk be

Tegyük fel, hogy van kezdőtőkéd, és olyan technológiába akarsz fektetni, amely öt év múlva valós piaci problémára ad választ. Szerintem ezeket az irányokat érdemes megnézni:

- **Adatproveniencia és validáció.** Erről szól az adat hitelességéről szóló fejezet. Ahogy a gép által generált tartalom elárasztja a netet, az ember által validált és tanúsított adat értéke exponenciálisan nő.
- **AI-audit és magyarázhatóság.** A szabályozók már most elvárják, hogy az AI-döntések magyarázhatók legyenek. Öt év múlva ez kötelező lesz a pénzügyi, orvosi és jogi területen.
- **Energiahatékony AI-infrastruktúra.** A nagy modellek energiaigénye fenntarthatatlan. A kis, specializált, helyben futó modellek (edge AI) piaca nagy lesz.
- **Digitális ikrek infrastruktúrához.** Hidak, csőhálózatok, épületek valós idejű digitális mása. Aki előre jelzi a meghibásodást, életet és pénzt takarít meg.
- **Kiberbiztonság az AI ellen.** Az AI ugyanolyan jól ír támadókódot, mint védelmi kódot: deepfake, automatizált social engineering, infrastruktúra elleni támadás. A védekezés ma gyerekcipőben jár.
- **Az ember és az AI együttműködésének eszközei.** Erről szól a strukturált kontextusról szóló fejezet. Minden iparágban szükség lesz rá, és ma még senki nem oldotta meg jól.
- **A szabályozási megfelelés automatizálása.** GDPR, AI Act, ágazati szabályozások: egyre több van belőlük, és egyre bonyolultabbak. Nagy tanácsadási területek nyílnak meg.
- **Az oktatás újragondolása.** Nem e-learningről beszélek, hanem arról, mit kell tudnia az embernek, ha az AI mindent tud. A kritikai gondolkodást, az ítélőképességet és az AI kimenetének ellenőrzését tanítani kell, és még senki sem tudja, hogyan.

Ha pisztolyt tartanának a fejemhez, és kettőt kellene választanom, az adatprovenienciát és az ember és az AI együttműködését mondanám. Az elsőt azért, mert minden más erre épül, a másodikat azért, mert ez a legközelebbi valós piac.

Spekuláció: saját kitekintés 2036-ra

Tíz év múlva a szellemi munka valószínűleg felismerhetetlenül átalakul. Nem tűnik el, de egyre specializáltabb lesz.

Lesznek szakemberek, akik üzleti, orvosi vagy jogi tudást strukturálnak a gépek számára, mások ebből a strukturált tudásból rendszereket építenek. Megjelennek olyan szerepkörök is, amelyeknek az AI-rendszerek validálása, auditálása és a döntésekért viselt felelősség kezelése lesz a fő feladata. A hagyományos szakmák sem tűnnek el, de várhatóan szűkebb, nagyobb hozzáadott értékű területekre húzódnak vissza.

A „programozó", az „ügyvéd", az „orvos" vagy a „könyvelő" mint általános foglalkozásnév valószínűleg ugyanolyan pontatlan lesz, mint ma az „informatikus": egy címke, amely többet elrejt, mint amennyit elárul.

Az adat valóban lehet az új olaj, de nem nyers formában. Az olajiparban a finomítás teremtette meg az igazi értéket, és az adatoknál is a strukturált, hitelesített és jogilag rendezett adatkészletek lesznek a legfontosabb erőforrások. Valószínűleg azok járnak jól, akik ezt a „finomítást" el tudják végezni.

A szabályozás utoléri a technológiát, de valószínűleg csak részben. Az AI Act és utódai keretet adnak, a részletszabályok viszont fokozatosan, iparági önszabályozásból, szakmai gyakorlatból és bírósági precedensekből állnak össze.

A szolgáltatóipar közben erősen átalakul. A személyes jelenlét egyre ritkább lesz, és sok területen az AI-asszisztens lesz az első kapcsolattartási pont. A bankfiók, az ügyfélszolgálati iroda és a biztosítási tanácsadó nem tűnik el teljesen, de kivétel lehet belőle.

Közben több kockázat is jól látható.

Az egyik a rejtett tudás. Ha az emberek már nem értik, miért jutott egy AI egy adott következtetésre, és a magyarázatot is egy másik AI-tól kérik, a tudás fokozatosan kicsúszhat az emberi ellenőrzés alól.

A másik a túl korai egyértelműsítés. Az AI hajlamos gyorsan feloldani a bizonytalanságot, és egyetlen megoldást tényként tálalni. Így könnyen elvesznek az alternatívák, a kivételek és azok a kompromisszumok, amelyeken sok szakmában a jó döntés múlik.

A harmadik a felelősség helye. Ha minden döntésre az a válasz, hogy „az AI ezt javasolta", a felelősség fokozatosan elpárolog. Hosszú távon a társadalmaknak újra kell definiálniuk, ki és milyen feltételekkel felel egy AI által támogatott döntésért.

Komoly kockázat a platformmonopólium is. Azok a vállalatok, amelyek egyszerre birtokolják a felhasználó adatait, az AI-asszisztenst és a szolgáltatási infrastruktúrát, soha nem látott befolyásra tehetnek szert. Egy látszólag „barátságos" digitális asszisztens könnyen kereskedelmi vagy politikai érdekek által befolyásolt kapuőr lehet.

Valószínűleg a digitális szakadék is mélyül. Akik nem férnek hozzá ezekhez az eszközökhöz, nem bíznak bennük, vagy nem tudják jól használni őket, egyre távolabb kerülhetnek a gazdasági és intézményi élet közepétől.

Mi marad emberi?

Valószínűleg a felelősségvállalás, a kontextus mély megértése, a kompromisszumkeresés, a bizalomépítés és a szakmai etika: az a képesség, hogy hiányos információ mellett is döntünk, és vállaljuk a következményeit.

Nem feltétlenül azért, mert az AI soha nem tud majd ezekben segíteni, hanem mert ezek alapvetően társadalmi és erkölcsi kérdések, és egy társadalomban a felelősséget végül mindig emberek viselik egymás felé.

Összegzés

A tanulmányban négy állítást fogalmaztam meg.

Először. Az adat hitelessége tíz éven belül jogi és piaci követelmény lesz. A technikai alapok már megvannak a C2PA-szabványban, de hiányzik a validátorok közössége, a jogdíj-elosztás mechanizmusa és a biometrikus emberi hitelesítés rétege. Ez etikai kérdés és üzleti lehetőség is.

Másodszor. Az AI sokszor nem azért hoz rossz döntést, mert rosszul következtet, hanem mert hiányos információból dolgozik. A megoldás ezért nem egyszerűen a „jobb AI", hanem a strukturáltabb információátadás: réteges specifikációk, élő dokumentumok, és kifejezett mechanizmusok a bizonytalanság és a hiányok kezelésére.

Harmadszor. Bizonyos szabályozott szakmákban az AI-asszisztált döntéseknek évekkel később is reprodukálhatónak kell maradniuk. Ehhez külön infrastruktúra kell: befagyasztott verziójú modellek, tartalomcímzett tárolás és auditálható döntési folyamatok.

Negyedszer. A vállalati, on-premise AI-rendszerek biztonsági problémái (az adatizoláció, a prompt injection elleni védelem és a szabályozási megfelelés) újfajta szakértelmet és új tanúsítási modelleket igényelnek. Ezt a területet ma sem a klasszikus IT-biztonság, sem a governance-re összpontosító nagy tanácsadó cégek nem fedik le teljesen.

A javaslataim nem kész megoldások, hanem lehetséges irányok. Ha öt év múlva visszanézek erre az írásra, és kiderül, hogy valamelyik feltevésem téves volt, az nem baj. Az lenne a hiba, ha ezekről a kérdésekről nem kezdenénk el időben gondolkodni.

A fejlesztői, mérnöki, orvosi és jogi munka átalakulása nem jelenti azt, hogy ezek a szakmák megszűnnek. Minden tudásalapú hivatásban elsősorban a szabálykövető, ismétlődő tevékenységek szorulnak vissza. Felértékelődik a kontextus megértése, a felelősségvállalás, a rendszertervezés, a validáció és a kompromisszumok kezelése.

A szolgáltatóipar személyes jelenléte valószínűleg erősen visszahúzódik, és a helyét egyre inkább digitális asszisztensek és AI-alapú ügyfélkapcsolat veszi át. Ez komoly etikai és társadalmi kérdéseket vet fel, és ezekre még azelőtt választ kell találnunk, hogy egy konkrét ember egy konkrét élethelyzetben szembesülne azzal, hogy a korábban természetes emberi segítség már nem érhető el számára.

Hivatkozások és további olvasmányok

- C2PA (Coalition for Content Provenance and Authenticity). Nyílt szabvány. https://c2pa.org
- IEEE 29148:2018. Systems and software engineering — Life cycle processes — Requirements engineering. Az ipari szabványos követelmény-leírás guideline-jai.
- ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. AI-menedzsment tanúsítható szabvány.
- ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management.
- NIST AI Risk Management Framework (AI RMF 1.0), 2023, és Generative AI Profile, 2024.
- OWL 2 Web Ontology Language. W3C ajánlás. https://www.w3.org/TR/owl2-overview/
- RDF 1.1 Concepts and Abstract Syntax. W3C ajánlás. https://www.w3.org/TR/rdf11-concepts
- DMN (Decision Model and Notation). Object Management Group. https://www.omg.org/dmn/
- Open Policy Agent. Nyílt forráskódú policy-as-code keretrendszer. https://www.openpolicyagent.org
- OWASP Top 10 for LLM Applications. https://owasp.org/
- MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems). https://atlas.mitre.org/
- EU AI Act. Az Európai Unió mesterséges intelligencia rendelete (2024).
- PSD2 (Payment Services Directive 2). Európai uniós irányelv, amely megnyitotta a banki adatokat versenytársak előtt; ez az open banking alapja.
- Anthropic Claude for Excel; Skills funkció. Az Anthropic alkalmazás-integrációi 2025-től.
- Straits Research. Global AI Training Data Market Report, 2025–2033.
- Grand View Research. AI Training Data Market Analysis.

A tanulmány gondolatainak nagy része az elmúlt másfél év hosszú beszélgetéseiből született. A beszélgetőtársak technológiai kollégák, akik nem kérték a nevük feltüntetését; az állítások, a koncepciók és a javaslatok a szerzőéi.