

A természetes nyelvi feldolgozóeszközök kis jogi irodákban történő használatának akadályairól és lehetőségeiről



Absztrakt: A számítógépes nyelvészet (NLP) által szolgáltatott átütő technológiák kihívást jelentenek a jogi irodák számára. Ez különösen igaz a kis jogi irodákra. A fejezet első részében megvizsgáljuk ezeket a technológiákat és azok alkalmazásait a jog területén. A fejezet második részében azokat a technikai és szervezeti akadályokat tekintjük át, amelyekkel a kis ügyvédi irodák az adaptáció (vagy akár annak hiánya) során szembesülnek. Végül összefoglaljuk az új innovatív technológiák által kínált technikai és szervezeti lehetőségeket a kis jogi irodák számára.

Tárgyszavak: kis jogi irodák; jogi informatika; mesterséges intelligencia; számítógépes nyelvészet (NLP); kis nyelvek; dokumentumosztályozás; entitásfelismerés; beszédfelismerés; beszédszintézis; gépi fordítás; hanguletelelemzés; anonimizálás; intelligens keresés; transzferatanulás; jogi ügymenedzsment szoftver; jogiforrás-kutató szoftver; e-Discovery; jogi csevegőrobot; kivonatolás; alternatív jogi tanácsadás

¹ Vadász Pál a Nemzeti Közszerológati Egyetem Információs Társadalom Kutatóintézetének vezető kutatója. Diplomáját az Eötvös Loránd Tudományegyetem Természettudományi Karán, PhD-jét az NKE Hadtudományi és Honvédtisztképző Karán szerezte katonai műszaki szakon. 1980 óta dolgozik főállásban az információtechnológia területén. Fő kutatási területe a jogi informatika, valamint a nemzetbiztonsági szolgáltatók és rendvédelmi szervek számonkérhetősége. Rendszeresen ad elő hazai és nemzetközi konferenciákon, valamint publikál könyvfejezeteket és tanulmányokat hazai és külföldi kötetekben, folyóiratokban.

² Görög György több szabadalommal rendelkező szoftverfejlesztő mérnök, PhD. Egészségügyi szoftverek (várólista-döntéstámogató, klinikai laboratóriumi szoftverek), majd jogi keresőrendszerek fejlesztésével foglalkozott. Mostani specializációja a jogi szövegek és más nyelvi struktúrák számítógépes elemzése, feldolgozása. A Montana Tudásmenedzsment Kft.-nél, valamint a Nemzeti Adatvagyon Ügynökség tanácsadójaként dolgozik.

³ Üveges István számítógépes nyelvész, a Szegedi Tudományegyetem nyelvtudományi PhD-iskolájának hallgatója, a Montana Tudásmenedzsment Kft. számítógépes nyelvész kutatója, a TK-MILAB szentimentelemzéssel foglalkozó részprojektjének résztvevője. Érdeklődési területébe tartoznak a szentimentelemzési feladatok, a természetesnyelv-feldolgozás, Plain Language Movement, valamint a mesterséges intelligencia és a gépi tanulás.

Bevezetés

A feladat

A jelen írás felhasználja a szerzők tanulmányát,⁴ amely az EU finanszírozásával az Európai Ügyvédi Kamarák és Jogi Társaságok Tanácsa (CCBE) és az Európai Jogászok Alapítványa (ELF) részére készült 2021-ben. Az eredeti tanulmány a kis jogi irodák lehetőségeit és a rájuk leselkedő veszélyeket vizsgálta az átütő hatású technológiák,⁵ a mesterséges intelligencia (*artificial intelligence* – AI) és különösen a számítógépes nyelvészet (*natural language processing* – NLP) fényében.

Kis jogi irodák (a továbbiakban KJI-k) alatt értünk minden olyan, jogi munkával foglalkozó nem állami szervezetet, amely kevesebb mint tíz főt alkalmaz. Ilyenek az ügyvédi és közjegyzői irodák.

A fejezetben válaszokat keresünk arra, hogy melyek azok a jelentős technológiai fejlesztések az NLP területén, amelyek befolyásolhatják a KJI-k jövőjét, valamint milyen technológiai és emberi vagy szervezeti⁶ akadályok gátolják e technológiák felhasználását. Igyekszünk rávilágítani a lehetőségekre ebben az új környezetben.

A vizsgálódás korlátai

Ez a fejezet nem a CCBE és az ELF véleményét tükrözi, már csak azért sem, mert a CCBE és ELF által publikált tanulmány sem tükrözi mindenben a szerzőkét.

A jogi szakirodalomban alig találtunk olyan tanulmányokat, amelyek a KJI-k fejlődését befolyásoló akadályokat és lehetőségeket vennék számba. A legtöbb tanulmány a nagy ügyvédi irodákra és nagyvállalati jogi osztályokra összpontosít, vagy egyáltalán nem tesz különbséget a jogi iroda mérete szerint.

A technológia olyan gyorsan változik, hogy nem az a kérdés, hogy egy idevonatkozó megállapítás elavul-e, hanem az, hogy mikor (valószínűleg hamarosan). Az új technológiák, módszerek és termékek gombamód szaporodnak. A helyzet és az ezt követő stratégia rendszeres újragondolást igényel.

Bár az eredeti feladat a kis jogi irodákra vonatkozott, fontosnak láttunk egy másik tényezőt is, amely befolyásolja többek között a jogi irodák helyzetét, és ez a *kis nyelvek* problémája. Amint az alábbiakban is látható, a kis népességű nyelvekkel

⁴ János Pál Vadász – György Görög – Csaba Csáki – István Üveges: *Opportunities and Barriers in the Use of Natural Language Processing Tools in the SME Law Practices*. JUST-JACC-EJU-AG-2019, grant 881527.

⁵ Az angolban elterjedt *disruptive technology* kifejezést használjuk konzekvensen átütő technológiaként.

⁶ Szervezeti alatt értjük a szociológiai, politikai, gazdasági, pénzügyi, jogi stb. körülményeket, amelyek befolyásolják az erőfeszítéseket és a folyamatokat.

dolgozó jogi irodák és jogi osztályok komoly hátrányban vannak a domináns nyelvek használóihoz képest. A magyar nyelv nyilvánvalóan a kis nyelvek közé tartozik.

Jelen fejezet nagy része íróasztal-kutatáson, valamint tudományos, üzleti és jogi szakértőkkel folytatott kommunikáción alapszik. Üzleti és tudományos publikációkat használtunk referenciaként a megfelelő tudás megszerzéséhez. Sajnos a Covid-19-járvány miatt nem volt lehetőség helyszíni konzultációkra. Következtetéseket vontunk le az NLP-technológiákat alkalmazó egyéb szakmai ágazatok adataiból is, mint például a vezetési tanácsadás, az információs és kommunikációs technológia (a továbbiakban IKT), a marketing, az egészségügy stb.

Számos piackutatást és jelentést tanulmányoztunk át, így például a következő szervezetekét: HSBC, Deloitte, KPMG, Altman Weil, Consero, Thomson Reuters, Aderant, The Law Society of England and Wales és az ABA (Clio által támogatott). Bár a mintavételük, az értékeléseik és a módszereik eltérőek, a jelen tanulmány következtetéseire nem tartottuk szükségesnek ezeket külön értékelni.

Mivel a tudományos publikációk nyelve az angol, a legtöbb forrás a mi témánkban is angol nyelven olvasható. A tudományos folyóiratok adatbázisaiban, például a Google Scholarban található jogi szakirodalmi kiadványok többsége is angol nyelvű forrásból származik, ami elfogultságot mutathat az EU tudományos forrásai vagy az angoltól eltérő nyelveken írottak hátrányára.

Köszönet

Köszönetet mondunk a Wolters Kluwer Germanynek és a Wolters Kluwer Hungarynek a 2020-as és 2021-es felméréseik grafikonjaiért, amelyeket Megyeri Andrea által feldolgozott formátumában vettünk át szerkesztésre.⁷

A jogi munkában alkalmazott NLP-alapú technológiák áttekintése

A jogi munkában alkalmazott NLP-alapú technológiákat ebben a részben két csoportra bontva tárgyaljuk. Elsőként bizonyos *NLP-technológiákat* ismertetünk, amelyeket a jogi munka területén használt szoftverekben is alkalmaznak. Másodikként NLP-t alkalmazó *legaltech-alkalmazásokat* mutatunk be, vagyis olyan termékeket és szolgáltatásokat, amelyek NLP-t használnak, és amelyeket KJI-k is működtetnek a mindennapi munkájukban.

⁷ The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report; 2020 Wolters Kluwer Future Ready Lawyer: Performance Drivers and Change in the Legal Sector – Technology Is at the Centre of Top Performance. *Wolters Kluwer*, 2020. július 16.

A felsorolás nem teljes körű, de igyekszünk egyrészt azokra összpontosítani, amelyek valószínűleg használnak (majd) a KJI-k, másrészt azokra, amelyek a legrelevánsabbak a következő részben ismertetendő akadályok és lehetőségek szempontjából.

A mesterséges intelligenciának számos definícióját ismerjük. Az intelligencia legfrappánsabb meghatározása: intelligencia, amelyet az intelligenciatesztel mérünk, a mesterséges intelligenciáé pedig az olyan intelligencia, amely nem emberi vagy állati. Más megközelítéssel olyan szoftver, amely a problémákat az emberhez hasonlóan oldja meg.

A klasszikus AI-alkalmazások szabályalapúak. Ez feltételekhez kötött komplex parancsrendszereket jelent (IF... THEN DO vagy WHEN... DO stb.). A korszerű AI már jóval kevésbé épít bináris jellegű szabályokra, sokkal inkább valószínűségekre, amelyeket bonyolult statisztikai módszerekkel dolgoz fel. Az NLP az AI azon részterülete, amely klasszikus nyelvészeti problémákat számítógépes programok segítségével old meg.

Általános NLP-technológiák, amelyeket jogi alkalmazásokban is felhasználnak

Az alábbi módszerekben az a közös, hogy mindegyikben központi szerepet kap a tudás reprezentációja. Ez teszi lehetővé, hogy az emberi tudás elérhetővé váljon a gépi tanulási algoritmusok számára, azaz hogy képes legyen az adatokból és a gépi tanulási megoldásokból származó tudást együtt felhasználni. Jó példa erre a tudásgráfok konstrukciója, amely szintén megpróbálja megragadni a fogalmak kapcsolatát (szemantikáját).

Dokumentumosztályozás

A dokumentumosztályozás olyan nyelvtechnológiai eljárás, amelynek célja a természetes nyelvi dokumentumokat tartalmuk alapján automatikusan, előre meghatározott kategóriákba (témakörökbe) sorolni. A besorolás történhet kézzel létrehozott szabályok (Boole-operátorokkal összekötött kulcsszókifejezések) segítségével vagy gépi tanulással is (a rendszer kézzel osztályozott példákban tanulva felállít egy témamodellt, amely alapján képes számára ismeretlen dokumentumok kategorizálására is).⁸

Entitásfelismerés

A névelem-felismerés vagy más néven entitásfelismerés (*named entity recognition* – NER) olyan nyelvi elemző eljárás, amelynek segítségével azonosíthatók egy

⁸ Dokumentum-osztályozás. *OpinHU Tartalomelemzés*.

szövegben a tulajdonnevek és a tulajdonnevekhez hasonlóan viselkedő (egyedi referenciával rendelkező) elemek, például dátumok, címek, telefonszámok, mennyiségek stb. Az azonosítás mellett elengedhetetlen a névelem típusának azonosítása is (például személynév, földrajzi név, szervezetnév). Ehhez gyakran szükség van a szövegkörnyezet figyelembevételével történő egyértelműsítésre is, például „az András utca 22. alatt lakom”; „tegnap András elfelejtette, hol lakik”.⁹

Szövegleiratozás, beszédfelismerés

A beszédfelismerés (*speech-to-text* – S2T) a számítógépes nyelvészet interdiszciplináris részterülete, amely olyan módszereket és technológiákat alkalmaz, amelyek lehetővé teszik a beszélt nyelv számítógépes felismerését és géppel olvasható szöveggé alakítását a hatékonyabb kereshetőség érdekében.

Beszédgenerálás, beszédsszintézis

A beszédgenerálás (*text-to-speech* – T2S) az AI-re épülő technológia, amelynek során a gépi úton olvasható, írott szöveget a beszédsszintetizátor érthető hangformátumra alakítja.

Gépi fordítás

A gépi fordítás olyan automatikus fordítási eljárás, amelynek során egy adott nyelvű szöveget egy másik nyelvre számítógépes program segítségével fordítanak le.¹⁰ A gépi fordítások hasznosságát nem kell bizonyítani a jogi munkában. Bár magyar–angol relációban a minőség megközelíti az ember által végzettét, a szöveget mindenképpen ellenőrizni kell. Gépi fordítóalkalmazások futhatnak a felhőben (Google Fordító, DeepL), vagy minősített szövegek kezelésére a külvilágtól szigorúan elzárt (offline, Faraday-kalitkában védett) környezetben (Systran).

Hangulatelemzés

A hangulatelemzés (*sentiment analysis*) a természetes nyelvi feldolgozás, a szövegelemzés, a számítógépes nyelvészet használata az affektív állapotok és szubjektív

⁹ Entitásfelismerés. OpinHU Tartalomelemzés.

¹⁰ Gépi fordítás. *Wikipédia*.

információk szisztematikus azonosítására, kinyerésére, számszerűsítésére és tanulmányozására.¹¹

Anonimizálás

A jogi alkalmazásokban igen gyakran van szükség a személyek azonosítását lehetővé tevő adatok eltüntetésére a dokumentumokból. Az anonimizálás a fentebb ismertetett technológiai megoldások közül többet is használ, hiszen önmagában a nevek eltüntetése egy dokumentumból nem szünteti meg feltétlenül az információk személyekhez köthetőségét. Az anonimizálás során szükség van más összefüggések felismerésére is.

Az úgynevezett szűkösség az anonimizálás fő ellensége. Sokan árulnak földimogyorót, így a személyes adatok eltávolítása esetén a földimogyoró-szerződés ügyfeleit nem könnyű azonosítani. Ha azonban az adott országban csak egyetlen társaság értékesít villamos energiát, az könnyen azonosítható magának a szerződésnek a tárgyából a szerződéses partner nevének ismerete nélkül is.

Ugyanakkor, ha túl sok konkrét részlet eltávolítanak, a dokumentum hasznossága súlyosan csökkenhet.

Keresés

A hagyományos, lexikális alapú keresés kimerül a kulcsszavak adott kombinációjához illeszkedő karakterek pontos sorrendjének keresésekor. Ez kizárja a találati listából az elgépett, illetve szinonim alakokat, és feleslegesen behozza az azonos alakú, de az adott összefüggésben más jelentéssel bíró szavakat. A karaktersor-alapú keresés eredménye lényegesen javítható szótövezéssel, stopszavak kiiktatásával stb. A szótövezés azonban elmoshatja a jogi értelmezés finomságait, például az „egészségügyi dolgozó” és az „egészségügyben dolgozó” különbségét.

A szemantikus keresésben a fő cél az, hogy a keresőkifejezésben szereplő szavak a statikus, szótári jelentésükön, illetve konkrét karaktersor-reprezentációjukon túl is értelmezhetők legyenek. A kulcs az, hogy megtaláljuk a kereső szándékait, a keresőkifejezés által megadott kontextust és az egyes szavak közötti kapcsolatokat. Az ilyen információk hozzáadásával relevánsabb, a kereső szándékának (szándéka-inak) jobban megfelelő eredményeket kaphatunk. Ilyen eszköz a kontrollált szótárak (szinonimaszótár, tezaurusz, taxonómia, ontológia) használata.

¹¹ Sentiment Analysis. *Wikipedia*.

Modern módszerek a kontextuskinyerésre és a tudásreprezentációra

A szövegben szereplő szavak, kifejezések egymáshoz való viszonyának felhasználásával mélyebb összefüggéseket lehet kiaknázni. Ennek eléréséhez új megközelítésekkel kellett integrálni. Újdonságuk a lexikális alapú kereséshez képest az volt, hogy a kifejlesztett modellek kontextuális információt tartalmaztak a szavak ábrázolásában. Ilyen a mélytanulás, amelynek során nagy méretű szövegtestekkel, úgynevezett tanítóadatokkal tanítanak modelleket új szövegek feldolgozása céljából. Egy egyszerű példa illusztrálja a módszert. Ha a „vár” kifejezés a „randi” kontextusában szerepel, akkor tudni lehet, hogy emberek találkozásáról van szó. Ha ugyanez az „ostrom” szövegösszefüggésben, akkor valamilyen erődítményt jelent.

2013-ban jelentek meg a szóbeágyazásos (*word-embedding*) módszerek. A beágyazásalapú megoldások jó példája az előre elsajátított nyelvi modellek vagy ismeretgráfok létrehozása. Az utóbbiak lényegében specializált ontológiák, amelyek inkább a gyakorlati alkalmazhatóságra összpontosítanak, mint a filozófiai mélységre. Nem kódolják a hétköznapi tudást, hanem sok entitásra összpontosítanak. A beágyazásoknak nehézséget okozhat a szinonimák és antonimák megkülönböztetése (hasonló szöveggörnyezetben); például az „ítélet” és „végzés” formailag azonos környezetben szerepelhetnek, de megkülönböztetésük – az alkalmazási cél függvényében – fontosnak bizonyulhat.

További fejlődést hozott a neuronhálózat-alapú megoldásokban a BERT (*Bidirectional Encoder Representations from Transformers*) 2018-as megjelenése. A BERT a Google által közzétett nyelvi modell. A BERT újdonsága abban rejlik, hogy kontextus alapján képes automatikusan megkülönböztetni egymástól az azonos formájú szavak jelentését. Az előző példát követve – „a fiú vár a lányra” és „ostrom alá vették a várat” – a „vár” szónak nyilván egészen más jelentése van. Hagyományos szóbeágyazások (mint a Word2Vec vagy a GloVe) ugyanazt a matematikai ábrázolást (jellemzően egy vektortér-ábrázolást) rendelik a „vár” szóhoz mindkét mondatban, míg a BERT különböző vektort generál a „vár” szóhoz a két mondatban, amelyből a jelentéskülönbség visszanyerhető. Így nincs szükség a kézi egyértelműsítésre (*disambiguation*), így ez a feldolgozást lényegesen egyszerűsíti. További előnye a BERT-nek, amely szubszavak, szócsónkok (nem nyelvtani szótagok) alapján dolgozik, hogy képes kezelni a modelltanulás során nem tapasztalt szavak jelenlétét, ami a gyakorlati alkalmazásokban sokkal erőteljesebbé teszi.

Az ilyen szemantikai különbségek rekonstruálása a keresőkifejezések feldolgozásában kulcsfontosságú a releváns eredmények visszanyeréséhez, különösen a jog esetében, ahol a közös szavak konkrét jogi kifejezéseként is előfordulhatnak.

Transzfertanulás

A transzfertanulás fontos technológia azon akadályok elkerülésére, amelyekkel a kis nyelvek (azaz szerényebb jogi szövegállománnyal, vagyis tanítóadattal, például kevesebb szerződés) szembesülhetnek. A transzfertanulás olyan eszköz, amely a tudást egy forrásból nyeri ki, és egy másik célbeállításra alkalmazza. Példa erre egy AI-modell (elő)tanítása egy adott nyelv általános szövegkorpuszán, majd a modell finomhangolása ugyanazon vagy egy másik nyelv speciális doménkorpuszán (például jogi szövegek). A transzfertanulásban az alaphálózat először egy alapadatkészletre és -feladatra van kiképezve, hogy képes legyen az adatok (esetünkben a szövegek) nagyon általános jellemzőinek megragadására. A finomhangolási fázisban az eredeti modellt (részben vagy egészben) használják kiindulópontként az adott feladatra kifejlesztendő új modellhez, ezért az első modell által kinyert eredeti „tudás” átkerül az utóbbi modellbe. A közelmúltban a Google BERT-eszközének és származékainak megjelenésével ez az eljárás dominanciára tett szert a területen.

Nyelvközi transzfertanulás

A keresés új perspektíváit nyitotta meg a nyelvközi transzfertanulás (*cross-language transfer learning*, a továbbiakban CLTL). Ebben a gépi tudás átkerül egy (erőforrásgazdag) forrásnyelvből egy másik (erőforrásszegény) célnyelvbe. A források ebben a tekintetben értelmező jegyzetekkel vagy nyelvtani, például szófaj- vagy mondatrészjelöléssel ellátott szövegek, valamint példák. Az egynyelvű és a többnyelvű megközelítés gyakran vegyes, még a képzési szövegek is különböző nyelvek lehetnek. Bizonyos NLP-modellek használata esetén a CLTL a jövőben alapvető technológiává válhat a kisebb nyelvek számára, amelyek korlátozott tanítóadatszövegállománnyal rendelkeznek.

A tudományos élet szereplői közt a BERT-modell a meghatározó. A BERT megszületése felvetette azt a reményt, hogy egyetlen eszköz megoldja az összes feladatot, legalábbis a GPT-3¹² megjelenéséig, például Ghavidel és szerzőtársai,¹³ valamint Wenjie¹⁴ szerint.

¹² Sejuti Das: GPT-3 Vs BERT For NLP Tasks. *Analytics India Magazin*, 2020. szeptember 11.

¹³ Hadi Ghavidel – Amal Zouaq – Michel Desmarais: Using BERT and XLNET for the Automatic Short Answer Grading Task. In *Proceedings of the 12th International Conference on Computer Supported Education*. I. CSEDU, 2020. 58–67.

¹⁴ Zhang Wenjie: *The Comparison between the Tools for Named Entity Recognition*. Master's Thesis. Auckland University of Technology, 2020.

Információkinyerés

A keresés új módszerei közé tartoznak a szimbólumalapú módszerek (relációkinyerés, információkinyerés), amelyek szervezik és értelmezik a jogi ismereteket és az egyes szimbólumok közötti kapcsolatokat (például nevesített entitások mint személyek, szervezetek). Számos alkalmazási lehetőséget tártak fel. Itt relációk alatt az egyes entitások (tárgyak, személyek, fogalmak, jellemzően főnevek) közötti kapcsolatokat értjük. Például a relációextrakció egy speciális típusú információextrakció (IE), ahol a fő cél egy strukturált adatbázis létrehozása a mondatokból a szövegből kivont entitáspárokkal és relációtípusokkal. Egy másik példa az események sorozatának felderítése egy büntetőügyben.

Dokumentumelemzés

A dokumentumelemzés szimbólumalapú megoldás, amely a szövegben található strukturálatlan információt olyan formába szervezi, amelyet emberek és szoftverek egyaránt értelmezhetnek. Ez úgy érhető el, hogy a szöveg entitásait kapcsolatba hozzák egymással, vagy ontológiákat, illetve szóhálókat építenek.

Jogi szoftverek (legaltech-alkalmazások) NLP-technológiákkal

A jogi ügymenedzsment szoftverek

A jogi ügymenedzsment (*legal practice management*) szoftver integrálja és automatizálja a jogi gyakorlatok *front-* és *back-office* tevékenységeit, beleértve az időnyilvántartást, a találkozók ütemezését, az esetkezelést, az üzenetküldést, a projektmenedzsmentet, a számlázást stb.

Jogi dokumentumkezelő szoftverek

A jogi dokumentumkezelő (*document management system* – DMS) szoftver megoldja a dokumentumok iktatását, tárolását, különböző szempontok szerinti osztályozását, kereshetőségét, selejtezését, a munkafolyamatok automatizálását és a hozzáférések naplózását.

Jogiforrás-kutató szoftverek

A jogiforrás-kutató (*legal research*) szoftverek és szolgáltatások összesítik a különböző forrásokból származó bírósági ítéleteket, jogszabályokat, rendeleteket és az esetjogot, valamint eszközöket biztosítanak a jogszabályok feltárásának és elemzésének automatizálásához. A jogi kutatás segíti a jogszabályok változásainak nyomon követését, a jogi vélemények kialakítását és támogatását, valamint a precedensek azonosítását.

E-Discovery szoftverek

Az *e-Discovery*¹⁵ a jogi, igazságügyi, vállalati stb. dokumentumhalmazokban adott tartalmú vagy típusú dokumentumok feltárását jelenti. Ilyen dokumentumok az e-mailek, a csevegések, az azonnali üzenetek, Word- vagy PDF-fájlok, a hang-/videófájlok és a közösségimédia-üzenetek. *E-Discovery* munkafolyamat például egy teljes vállalati szövegtörzsben a csalás- vagy kartellgyanús elemek felderítése,¹⁶ valamint a törvényi megfelelés vizsgálata (*compliance review*).

Az adatfeltárási munkafolyamat fázisai a következők: a releváns források azonosítása, adatrögzítés, adatfeldolgozás, felülvizsgálat és terjesztés. AI-t használnak a feldolgozási fázisban meghatározott adatelemek azonosítására (*technology assisted review* – TAR).

Bírói döntések előrejelzése

A bírói döntések előrejelzése (*litigation prediction*) a bírósági döntések hatalmas szövegtörzsain alapuló, prediktív elemzések alkalmazása. Segít annak eldöntésében, hogy egy adott ügyet bíróságra vigyünk-e, vagy megegyezzünk. Felméri a lehetséges siker valószínűségét a tényállás alapján összevetve hasonló tényállású ügyek kimenetelével. Feladata lehet az igazságügyi szakértőre történő javaslattevés vagy az ellenfél szakértőjének értékelése.

Jogi csevegőrobotok

A jogi csevegőrobotok (*lawbots*) NLP-alapú válaszadó (*human-machine interfaced question and answer* – Q&A) alkalmazások, amelyek emberek közötti beszélgetést

¹⁵ Ennek a fogalomnak nem honosodott meg magyar megfelelője.

¹⁶ Pál Vadász – András Benczúr – Géza Füzesi – Sándor Munk: Identification Illegal Cartel Activities from Open Sources. In Babak Akhgar – Petra Saskia Bayerl – Fraser Sampson (szerk.): *Open Source Intelligence Investigation*. Cham. Springer, (h. n.), 2016. 251–273.

szimulálnak.¹⁷ A csevegőrobotok megértik a természetes nyelven hangban vagy írásban feltett kérdéseket (lásd *Szövegleiratozás, beszédfelismerés*), és megválaszolják azokat.

Kivonatolás

A kivonatolás (*summarisation*) során az AI-alkalmazás egy hosszabb szövegből kiszűri a lényegét. A folyamat kétféleképpen történhet. Vagy kiemel fontosnak tartott szövegrészeket (mondatokat) változatlan formában, vagy szabadszövegű összefoglalót készít.

Akadályok az NLP alkalmazásának útjában a kis jogi irodákban

Alapvetően kétféle akadály gátolja az NLP KJI-k általi befogadását. Egyeseket technológiai jellegűnek lehet minősíteni, a többinek pedig inkább az emberi természethez vagy a szervezeti jellemzőkhöz van köze.

Technikai akadályok

A tanítóadatok elérhetősége

Ha a tanítóadatok mennyisége nem elegendő a modellépítéshez, nem kapunk kielégítő eredményt. Ez a probléma olyan szakmai területeken vagy nyelveknél merülhet fel, amelyek esetében kevés dokumentum keletkezett, vagy kevés elérhető, akár jogi akadályok okán is. Erre példák lehetnek a személyiségi jogok miatt nem publikálható egészségügyi jelentések vagy éppen bírósági határozatok. Ilyenkor anonimizálják a dokumentumokat a további feldolgozhatóság végett (lásd *Anonimizálás*).

Kevés tanítóadat birtokában olykor transfertanítást alkalmaznak modellépítésre (lásd *Transzfertanítás*). A transfertanítás különleges esete egy nyelvi modell transzformálása egy másik nyelvre (lásd *Nyelvközi transfertanulás*).

Több központú nyelvek

Elméleti nyelvészeti szempontból a több központú nyelveket (*polycentric languages*) úgy lehet meghatározni, mint egynél több országban jelen lévő nyelveket, és mint az „alapnyelv” egyenértékű változatát (amely megegyezik az adott ország anyanyel-

¹⁷ Kate Brush – Jesse Scardina: Chatbot. Definition. *TechTarget*.

vével).¹⁸ Rudolf Muhr¹⁹ a több központú nyelveket a nyelvek és a dialektusok külön kategóriájának tekinti. Mivel közigazgatási és állami nyelvként vagy regionális hivatalos nyelvként működnek, nyelvi és kommunikációs autonómiával rendelkeznek. A német, angol, francia, görög, olasz, holland, portugál, spanyol, magyar és svéd nyelvet több központú európai nyelvnek tekintik az EU-ban.

Az ilyen nyelvek különös jelentősége a jogi szövegek területén az, hogy eltérő használatuk és hagyományaik miatt gyakran ugyanazt a jogi kifejezést alkalmazzák az azokat használó országok különböző lexikális kifejezései. Ennek egyik legikonikusabb példája a német, ahol például Ausztriában és Németországban különböző jogi fogalmakra hivatkoznak, például: *Jus* (Ausztria) vs. *Recht* (Németország) a jogra, *Ehepakt* (Ausztria) vs. *Ehevertrag* (Németország) a házassági szerződésre és az *Arbeitsbewilligung* (Ausztria) vs. *Arbeitserlaubnis* (Németország) munkavállalási engedélyre utal.²⁰

Az EU-n belül maradva, ez a kódrendszer a DEU-t mint „német (Németország)”, a DEA-t pedig mint „német (Ausztria)” listázza. Egyértelmű, hogy a Németországban és Ausztriában használt hivatalos írott nyelvek között is vannak különbségek.²¹ Az ilyen több központú nyelvek a modell tanításakor különleges figyelmet igényelhetnek.²² A különbségek azonban ugyanúgy tekinthetők nyelvi vagy jogtudományi különbségeknek. Mindkét esetben a transzfertanulás a nagyobb szövegkorpuszt használó mesterséges intelligencia úgynevezett előtanítását és a kisebb korpuszt használó finomhangolást jelenti. A gyakorlati folyamat nagyrészt ugyanaz lesz, és az eredmény két különböző módon képzett AI-modul lesz.

A jogi nyelv sajátosságai

A jogi nyelv és reprezentációja jelentősen eltérhet a mindennapi nyelvtől. A szabályozási szövegeket és bizonyos mértékig a szerződéseket is összetett mondatok jellemzik, beleértve a számozott listás hivatkozásokat stb. Sanchez az amerikai,²³

¹⁸ Pluricentric Language. *Wikipedia*.

¹⁹ Rudolf Muhr: Die plurizentrischen Sprachen Europas – ein Überblick. In Eva Gugenberger – Mechthild Blumberg (szerk.): *Europa. Zur Situation der regionalen Sprachen von der Iberischen Halbinsel bis zum Kaukasus*. Österreichisches Deutsch – Sprache der Gegenwart, II. Frankfurt am Main, Peter Lang Verlag, 2003. 191–233.

²⁰ Heidemarie Markhardt: Das Österreichische Deutsch im Rahmen der EU. Frankfurt am Main – Berlin – Bern – Bruxelles – New York – Oxford – Wien, Peter Lang, 2005.

²¹ Rudolf Muhr: The Differences in the Legal Terminology of Austria and Germany and the Results for the German Legal Terminology within the Scope of the European Union. *Muttersprache*, 119. (2009), 3. 199–216.

²² Rudolf Muhr – Josep-Àngel Mas Castells – Jack M. Rueter: *European Pluricentric Languages in Contact and Conflict*. New York, Peter Lang, 2020.

²³ George Sanchez: Sentence Boundary Detection in Legal Text. In *Proceedings of the Natural Language Processing Workshop 2019*. Association for Computational Linguistics, 2019. 31–38.

Glaser és munkatársai a német,²⁴ míg Vincze a magyar jogi szövegekről²⁵ közöl részleteket. A számozott (és betűvel írt) listák miatt a mondatok gyakran sorvégi karaktereket tartalmaznak; a mondatok egyes nyelveken belső mondatzáró pontot is tartalmazhatnak a sorszámok és rövidítések miatt. A bekezdések átléphetik az oldalhatárokat, ami bármilyen szövegtípusú optikai karakterfelismerés esetén gyakori probléma. Manapság a római számok nem túl gyakoriak a mindennapi nyelvben. A latin nyelv is előfordulhat a jogi dokumentumokban.

Szokás a jogalanyok, például pártok, rendeleti jellegű jogi aktusok stb. rövidítése vagy szokásos nevének használata úgy, hogy teljes nevük csak egyszer szerepeljen a szövegben („továbbiakban”). Az anonimizálás hármaspontokat (...) vagy teljes pontsorozatokat (....) hagyhat, és a {} és [] zárójeleknek, továbbá más speciális karaktereknek (például §►”) is van szerepe, nem beszélve a perjelek továbbra is előforduló használatáról a kerek zárójelek helyett.

Pozitívum viszont, hogy a jogi szövegekben csak kevés elírás valószínű, kivéve talán a nyers vázlatokat. A jogi nyelvezet morfológiai sokfélesége bizonyosan kisebb, mint a mindennapi beszéd, hiányzik az informális megszólítás (vagy egyáltalán a megszólítás), az első és második személy, és kevesebb az ige stb. A jogi szövegek ugyanakkor más (szabályozott vagy szerződött) szakmákra vonatkozó különleges feltételeket is tartalmaznak. Ez *de novo* tanítást igényel inkább, mint a hétköznapi nyelvből történő transzfertanítást.

A dokumentumok mérete és a mondatok hossza is jelentős tényezők a mesterséges intelligencia tanításával kapcsolatban.

A megoldás ezekre a problémákra – a mesterséges intelligencia tanítása előtt – a speciális szöveg-előkészítés és a neurális hálózat speciális konfigurálása lehet a dokumentum és a mondat méretének kezelésére.

Ugyanakkor azt is meg kell jegyezni, hogy az egyszerűbb neurális hálózati modellek is betaníthatók kevesebb szöveggel, és kielégítően működhetnek. A nagy adatkészletek iránti igényt csökkentő technológiákat Maheswari,²⁶ a későbbi fejlesztéseket például Riekert és munkatársai²⁷ tanulmányai ismertetik.

²⁴ Ingo Glaser – Sebastian Moser – Florian Matthes: Sentence Boundary Detection in German Legal Documents. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*. II. ICAART, 2021. 812–821.

²⁵ Vincze Veronika: A Miskolc Jogi Korpusz nyelvi jellemzői. In Szabó Miklós – Vinnai Edina (szerk.): *A törvény szavai*. Miskolc, Bíbor Kiadó, 2018. 9–36.

²⁶ Jyoti Prakash Maheswari: Breaking the Curse of Small Data Sets in Machine Learning: Part 2. *Medium*, 2019. április 27.

²⁷ Martin Riekert – Matthias Riekert – Achim Klein: Simple Baseline Machine Learning Text Classifiers for Small Datasets. *SN Computer Science*, 2. (2021). 178.

A fekete doboz és az ügyvédek felelőssége

A gépi mélytanulás (*deep learning*) úgynevezett fekete dobozos (nem átlátszó) neurális hálózatokat eredményez, amelyek nem tudják maguk megmagyarázni az eredményeiket. Még ha ezek az eredmények megbízhatók is, amint azt a kiterjedt tesztelés kimutatta, maga a folyamat statisztikai jellege sokszor kétségeket hagy. Nehéz elképzelni Európában, hogy egy KJI vagy egy bíróság normál terhelés mellett komoly jogi döntést hozzon pusztán azért, mert egy program azt mondta. Ez még az olyan automatikus kockázatértékelés (például annak valószínűsége, hogy az alperes visszaesővé válik) esetén is igaz, mint amelyet a COMPAS-rendszer készít az Egyesült Államokban,²⁸ vagy amelyet a holland alternatív vitarendezési (ADR) kezdeményezés, az e-Court²⁹ ajánlott. Jelenleg komoly erőfeszítések történnek az AI kimeneteinek megmagyarázhatóságával kapcsolatban. Például az Európai Unió AI fehér könyve szerint „a szimbolikus érvelés és a mély idegi hálózatok kombinálása segíthet a mesterséges intelligencia kimeneteinek érthetőségének javításában”.³⁰ További követelmény lehet, hogy a magyarázatnak természetes nyelven kell történnie (esetleg grafikonokkal és más vizualizációs eszközökkel alátámasztva).

Ami a jogi NLP-technológiákat illeti, kiváló és részletes felmérések állnak rendelkezésre a problémáról, amelyek potenciális megoldásokat kínálnak, bár úgy tűnik, hogy némelyikük kívül esik a KJI-k területén.³¹

Technológiai különbségek a kontinentális és a common law országok között

Mivel a tudományos és szakmai publikációk nagy része angolszász országokból származik, felvetődött a jogos kérdés, hogy a jelen fejezetben szereplő kijelentések hatással vannak-e a kontinentális jog gyakorlóira. A válasz egyszerű: tisztán számítástechnikai szempontból nincsenek jelentősebb következményei ennek a különbségnek, minden NLP-eszköz és -módszer alkalmazható mindkét tartományban. A *common law*-ban alapvető fontosságú, hogy minden vonatkozó bírósági ítélet rendelkezésre álljon, míg a kontinentális rendszerekben a kodifikált alapszabályok vannak túlsúlyban, de a konkrét NLP-technológiák (lásd *Jogi szoftverek...*) között nincs különbség.

²⁸ Andrew Lee Park: Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing. *UCLA Law Review*, 2019. február 19.

²⁹ Willemien Netjes – Arno R. Lodder: e-Court – Dutch Alternative Online Resolution of Debt Collection Claims. *International Journal of Online Dispute Resolution*, (2019), 1. 96.

³⁰ European Commission: *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. COM(2020)65 végleges (2020. február 19.).

³¹ Zsolt Zódi: Algorithmic Explainability and Legal Reasoning. *The Theory and Practice of Legislation*, 10. (2022), 1. 67–92.

A kis nyelvek mint technikai akadály

A gépi tanulási feladatok mindig nagy szövegtesteket igényelnek. Míg egy általános nyelvi modell felépíthető az adott nyelvhez tartozó Wikipédia-bejegyzésekből, egy szakterület (például jogi entitásfelismerő) különleges szövegtestből képzett tanítóadatot igényel. Ezeket sokszor kézzel kell címkézni, amihez szükség lehet nyelvészek és informatikai szakértők bevonására. Különböző módokon kell felkészíteni például egy korpuszt a hangulatelemzéshez vagy az anonimizáláshoz, és ehhez különbözőféle szakértelemre van szükség. Az ilyen kompetencia korlátozottabban érhető el a kis nyelvek esetében.

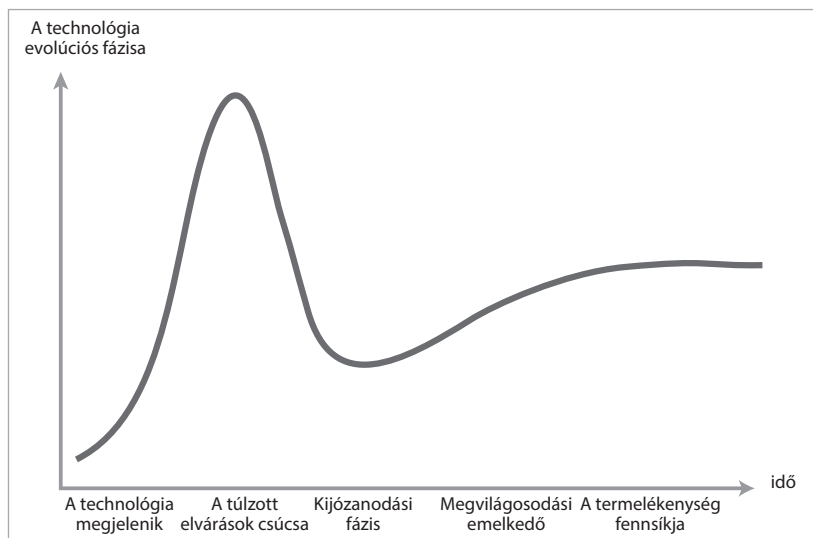
A kis nyelvek esetében, még ha a nyelvi modell felépítéséhez szükséges anyag online elérhető is, gyakran nem áll rendelkezésre olyan és annyi tanítóadat, amely és amennyi konkrétabb feladatokra is alkalmas lenne, vagy nem létezik olyan nyilvánosan elérhető tanítóadat, amellyel egy modell finomhangolása elvégezhető.

Ha a kisebb cégek egyéni igényeit is figyelembe vesszük, az általánosan elérhető források hiánya növeli a költségeket (minden vállalatnak magának kell megteremtenie a tanításhoz szükséges erőforrásokat), és a tanításhoz használt szövegtestek eltérő minősége nagyon eltérő minőséget eredményezhet.

Szervezeti akadályok

A Gartner-féle hype-ciklus

A technológiai fejlődés elfogadása, a technológiai megoldások tényleges alkalmazása, különösen az információs és kommunikációs iparban egy tipikus mintát, az úgynevezett *Gartner Hype Cycle*-t mutatja (1. ábra). Amint azt például a kézírás-felismerés, a dotcombuborék vagy az Y2000 fordulat esetében is láthattuk, a mesterséges intelligencia mint technológia elfogadottsága hasonló görbét mutat. A kezdeti lelkesedés, amely gyakran jellemzi a sajtót és a befektetői közösséget, eléri a csúcst, amikor a technológia korlátai kezdenek megmutatkozni. Ezt a kiábrándultság szakasza követi, amelyből a reális ipari növekedés sokkal később alakul ki, amely végül eléri a piac telítettségének szintjét. Az ábrán azonban nem látható, hogy a görbén lehetnek felfelé és lefelé irányuló ingadozások – ahogy az AI esetében történt a hetvenes-nyolcvanas évek fordulóján, valamint a kilencvenes évek elején – az úgynevezett AI-teleken (*AI-winter*). Mivel minden elemző egyetért abban, hogy az *NLP megítélése a legaltech területén jelenleg túlzottan optimista*, nagyon fontosnak tartjuk csillapítani a túlzott várakozásokat, és óvunk mindenkit a kellemetlen meglepetésektől.

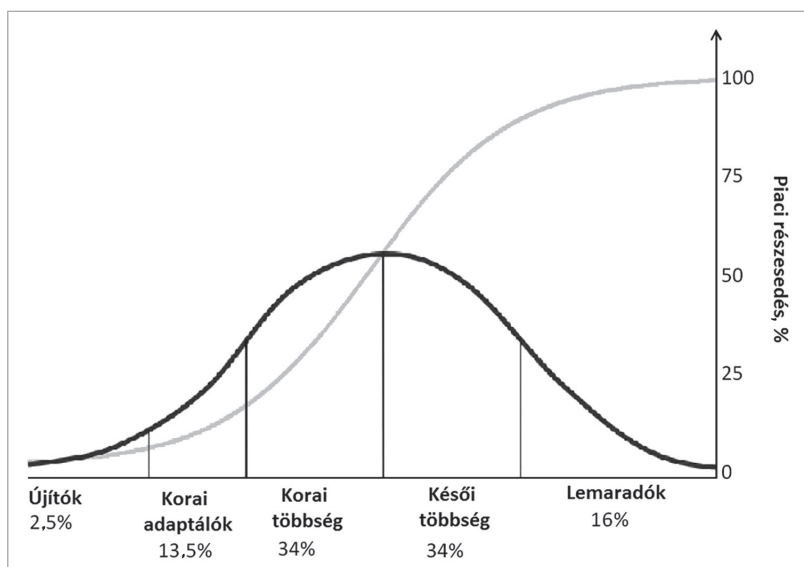


1. ábra: *Hype*

Forrás: Kasey Panetta: 5 Trends Appear on the Gartner Hype Cycle for Emerging Technologies, 2019. *Gartner*, 2019. augusztus 29.

A Gartner-féle *hype*-ciklus lazán kapcsolódik a Rogers-féle diffúziós görbéhez, amely megmutatja, hogyan, miért és milyen ütemben terjednek az ötletek és technológiák. A témát W. D. Henderson³² tárgyalja. Míg a Gartner-féle *hype* a várakozásokat mutatja, a Rogers-féle diffúziós görbe a különböző újítók piaci részesedését szemlélteti az idő függvényében. Ha feltételezzük, hogy a Rogers-féle diffúziós görbe nem egészen szimmetrikus haranggörbe, akkor azt is mondhatjuk, hogy a Gartner-görbe alatti terület fedi le a Rogers-görbe alatti újítót (2. ábra).

³² William D. Henderson: Innovation Diffusion in the Legal Industry. *Dickinson Law Review*, 122. (2018), 2. 458.



2. ábra: A Rogers-féle innovációs görbe

Forrás: Wikipédia ([wikimedia.org/wiki/File:Rogers_diffuzio_elmelete](https://commons.wikimedia.org/wiki/File:Rogers_diffuzio_elmelete))

3.2.2. A kis nyelvek mint szervezeti akadály

Az *Ethnologue* a kis nyelvet tízezernél kevesebb beszélővel definiálja.³³ Európai szövegösszefüggésben úgy értelmezik, hogy egy uniós nyelv kicsi, ha kevesebb mint tízmillióan beszélik. Még ha egy nyelvnek több mint tízmillió beszélője is van, a lakosság egyes esetekben több joghatóságra osztható, így a tényleges homogén (az itt tárgyalt értelemben) népesség még kisebb lehet – lásd a több központú nyelvekről írt alfejezetet. Ilyen – policentrikus – nyelv a német, a francia vagy éppen a magyar is.

Jelentős probléma, hogy a kis nyelvek, beleértve a kis többközpontú, policentrikus nyelveket (lásd *Többközpontú nyelvek*) is, *nem érik el a méretgazdaságosság küszöbét a gyártók és a szolgáltatók számára*. A piac túlságosan töredezett, túl sok kis résztvevőből áll ahhoz, hogy a kis nyelveken futó alkalmazásokba érdemes legyen befektetni. Más szavakkal: egy alkalmazás befektetésének megtérülése a nagyobb nyelv/lakosság/piac esetében sokkal magasabb, mint egy kisebb esetében. Ennek eredményeként a kis nyelvek alkalmazásainak választéka és minősége sokkal gyengébb, mint a nagyoké. Három példa alább.

- Egy gyors pillantás az angol és a lett Wikipédia cikkeinek számáról elég meggyőző (6 546 475 szemben a 114 648-cal).³⁴

³³ Size and Vitality of Shan. *Ethnologue*.

³⁴ Multilingual Statistics. *Wikipedia*.

- A Microsoft *Word's Style Guidance (/File/Options/Proofing/When Correcting Spelling and Grammar in Word/Settings)* eszközben látható az angol és a német lehetőségek bősége, a magyarok szűkössége, vagy akár a lett vagy katalán lehetőségek teljes hiánya (2021. augusztus 20-án).
- A piacvezető nyelvtan- és stílusellenőrző Grammarly nem is érhető el más nyelven, csak angolul.³⁵

Az UNESCO öt szinten határozza meg egy nyelv veszélyeztetettségét (a „biztonságos” után) az alábbiak³⁶ szerint:

- Sérülékeny: „a legtöbb gyermek beszéli a nyelvet, de ez bizonyos területekre (pl. otthon) korlátozódhat”;
- Erősen veszélyeztetett: „a gyermekek már nem anyanyelvként tanulják a nyelvet az otthonukban”;
- Súlyosan veszélyeztetett: „a nyelvet a nagyszülők és az idősebb generációk beszélik; bár a szülőgeneráció megértheti, nem beszélik a gyerekekkel vagy egymás között”;
- Kritikusan veszélyeztetett: „a legfiatalabb beszélők a nagyszülők és még idősebbek, ők is csak olykor, ritkán beszélik a nyelvet”;
- Kihalt: „már nem beszéli senki; az 1950-es évek óta feltehetően kihalt.”

Egy adott nyelv *digitális jelenléte* különböző tényezőktől függ, mint például a közösség mérete, presztízse, pénzügyi háttere, iskolázottsági szintje, a politikai lobbihatalma, technológiai kifinomultsága stb. Egy nyelvnek lehet viszonylag kicsi a népessége, ugyanakkor jelentős vagy éppen nulla a digitális jelenléte. Az afrikai mandinga (1,35 millió beszélő) írástudóinak aránya 1% alatt van, a Wikipédia-jelenléte pedig zéró. A romani (cigány) nyelvet Európában 1,5 millió ember beszéli, és nincs Wikipédia-jelenléte (bár a dél-európai Vlach Romani változatban 690 cikk található). A piemonti nyelvet 2-3 millióan beszélték Torino régióban, és nincs jelentős digitális jelenléte. Ugyanakkor az észt 1,165 millió beszélővel, 221 ezer Wikipédia-szócikkkel rendelkezik. A 2016-ban beszélt³⁷ 7139 nyelv közül 321-nek van legalább 1 cikke a Wikipédiában.³⁸ Feltételezhető tehát, hogy a világ nyelveinek 4,5%-a rendelkezik valamilyen digitális jelenléttel. A *Digital Language Death*³⁹ arra a következtetésre jut, hogy a nagyjából 7000 még élő nyelv közül 2500 még egy évszázadig fennmaradhat, és a *digitális túlélők becsült száma világszerte mindössze 250*.

³⁵ Robert Dale – Jette Viethen: The Automated Writing Assistance Landscape in 2021. *Natural Language Engineering*, 27. (2021), 4. 511–518.; Does Grammarly Support Languages Other than English? *Grammarly Support*.

³⁶ List of Endangered Languages in Europe. *Wikipedia*.

³⁷ How Many Languages Are There in the World? *Ethnologue*.

³⁸ List of Wikipedias. *Wikimedia*.

³⁹ Kornai, András: Digital Language Death. *PLOS ONE*, 2013. október 22.

Az Európai Unióban 70 beszélt és 24 hivatalos nyelv van.⁴⁰ Ezen nyelvek némelyikét egy kis közösség gyakorolja. Nem találtunk adatokat mind a 70 nyelv digitális jelenlétéről, nem is beszélve az NLP-használatuk kifinomultságáról. A digitális jelenlét ugyanakkor létfontosságú a számítógépes nyelvészetre épülő alkalmazások elterjedésének szempontjából, hiszen ennek alapvető feltétele a megfelelő mennyiségű és minőségű tanítóadat elérhetősége. Más szavakkal: *megfelelő digitális jelenlét nélkül egyes szakmák nem tudnak élni a nyelvükben, és más nyelvben fognak tovább élni.*

Alternatív jogi tanácsadás

Az alternatív jogi tanácsadói szolgáltatások (*alternative legal service provider – ALSP*) kvázi jogi szolgáltatásokat nyújtanak magas szintű NLP-alapú technikai arzenállal az interneten.⁴¹ Miután beruháztak és szakosodtak ezekre az átütő hatású technológiákra, rendkívül versenyképes árszínvonalon tudnak működni. Az ALSP-robotok anélkül válaszolhatnak az ügyfelek által gyakran feltett kérdésekre, hogy azoknak jogi irodához kellene fordulniuk. Az Egyesült Államok jogi piacán megfigyelhető az a tendencia, hogy egyes vállalati jogi osztályok jobban bíznak az ALSP-kben, mint a jogi cégekben.⁴² Ez azt is jelenti, hogy az ALSP-k részt hasítanak ki a jogi irodák piacából, azaz bizonyos területeken fenyegetést vagy legalábbis versenyt jelentenek a klasszikus jogi irodáknak.⁴³

A csevegőrobot mint az emberi tanácsadás lehetséges alternatívája

A csevegőrobotok (lásd *Jogi csevegőrobotok*) nemcsak jelentősen olcsóbbak, mint az emberi munkaerő, hanem korlátozás nélkül rendelkezésre állnak, valamint a megszerzett információkat törlésükig tárolják, ellentétben az emberekkel, akik magukkal viszik őket, amikor távoznak, így az újonnan érkezőket újra kell képezni. A csevegőrobotok minősége folyamatosan fejlődik. Ugyanakkor kétségtelen, hogy minél bonyolultabb a probléma, annál valószínűbb, hogy a kérést továbbítani kell az embernek. Egy bizonyos pontig a csevegőrobotok állják a Turing-tesztet.⁴⁴

A Gartner 2019-es előrejelzése szerint a jogi csevegőrobotok, azaz AI által működtetett virtuális jogi asszisztensek 2023-ban a jogi megkeresések egyharmadát

⁴⁰ Languages of Europe. *Wikipedia*.

⁴¹ What is an Alternative Legal Service Provider (ALSP)? [ParalegalEDU.org](https://www.paralegal.edu.org/)

⁴² Frank Ready: Legal Departments Likely Trust ALSPs More than Firms on Tech Recommendations. [Law.com](https://www.law.com), 2021. augusztus 4.

⁴³ Rachel Zahorsky – William D. Henderson: Who's Eating Law Firms' Lunch? *ABA Journal*, 2013. október 1.

⁴⁴ Turing-teszt. Wikipédia.

fogják kezelni.⁴⁵ Még ha ez a jóslat túl optimista/pesszimista is, a csevegőrobotok egyre nagyobb részt követelnek az ügyfélszolgálati piaci szegmensből, és nincs okunk feltételezni, hogy a jogi területen mást tapasztalnánk.⁴⁶

Kis és nagyobb cégek

A KJI-k elveszíthetik-e piaci részesedésüket a nagyobb jogi irodákkal szemben? Mivel a fejlett technológia beruházásigényes, a kisebb szervezetek nem engedhetik meg maguknak a csúcskategóriás megoldásokat, így az alacsonyabb hatékonyság okán piacot veszíthetnek.

A piac mérete

A jogi piacon, ahogy sok más iparágban is, a „méret számít”: a *Deutscher Anwalt Verein* számára készített tanulmány szerint 2030-ra a jogi irodák fennmaradása főleg az iroda méretétől függ majd.⁴⁷ A tanulmány nem mélyed el a témában, és nem közöl becsléseket a túlélési arányról, de általánosságban hangsúlyozza a digitális átalakulás tendenciáihoz való alkalmazkodás fontosságát. A LexisNexis 2019. évi Bellwether-jelentése⁴⁸ számos olyan kihívást említ, amelyekkel a KJI-knek szembe kell nézniük. Ezek egyike ugyancsak a növekedési korlát. A tanulmány mindazonáltal optimista, és még a technológiát sem említi kulcsfontosságú tényezőként, nemhogy az NLP-t.

Átállás jobban támogatott nyelvi környezetekbe

Az ügyfelek, illetve velük közösen a jogi irodák dönthetnek a hatékonyság okán úgy, hogy egy másik nyelven, esetleg – a legjobban támogatott *lingua franca* – brit angol vagy még inkább amerikai *angol nyelven dolgoznak, ha egy megfelelő NLP-arzenál nem érhető el az anyanyelvi környezetben*. Ez természetesen csak olyan esetekben lehetséges, amikor a bíróság, a joghatóság és a nyelv választható.

⁴⁵ Gartner Predicts That by 2023 “Lawbots” Will Handle a Quarter of Internal Legal Requests. *Gartner*, 2019. január 30.

⁴⁶ Kayla Matthews: Applications for Legal Chatbots. *Law Technology Today*, 2020. július 13.

⁴⁷ Executive Summary der Rechtsdienstleistungsmarkt 2030. Deutscher Anwalt Verein – Prognos. (H. n.), 19.

⁴⁸ The Bellwether Report Series 2019: Is the Future Small? *LexisNexis*.

Gyenge beszállítói támogatás

Ha egy KJI külső beszállítót választ, többféle kihívással is szembesülhet. Egyrészt a túl lelkes, induló vállalkozások sokszor nem képesek megfelelő támogatást nyújtani, vagy akár el is tűnhetnek. Másrészt az érett, szilárd vállalatok árazása nehezen kitermelhető, vagy a KJI-k számára egyenesen elviselhetetlen.

Az EU hátrányos helyzetben van az USA-val vagy Kínával szemben

Az EU országai több szempontból is hátrányos helyzetben vannak az USA-hoz és Kínához képest. Az EU piaca különböző joghatóságok szerint oszlik meg, és a különböző joghatóságok között lehetnek *protektionista, mesterséges vagy bürokratikus akadályok*.

Minden előnye ellenére a GDPR negatív hatással van az AI-alapú induló vállalkozásokba történő kockázatitőke-befektetésekre.⁴⁹

A fő akadály azonban a *nyelvi megosztottság*. Ez a piac széttagozottságát okozza, mivel az egyik országban működő alkalmazások nem használhatók egy másikban. Nyelvenként *viszonylag csekély mennyiségű tanítóadat* áll rendelkezésre szemben a nagy, homogén piacokkal (USA, Kína). Ennek eredményeként a nagy beszállítók sem érdekeltek abban, hogy termékeiket kis nyelvi piacokra lokalizálják.

Működési költségek

A szakemberhiányon túl fontos megemlíteni, hogy a szoftveralkalmazások bevezetésének és támogatásának költségei általában magasak. Ezek a költségek arányosan könnyebben teríthetők a nagyobb jogi irodáknál, amelyek nagyobb tranzakciókat bonyolítanak, mint a KJI-k. Ezen túl az egyébként könnyen automatizálható kisebb jogi feladatok száma nem éri el a *méretgazdaságosság* szintjét a KJI-k esetében.

A költségek és a hatékonyság

A jogi szektort – és különösen a KJI-ket – fenyegető legnagyobb veszély az, hogy azokat a feladatokat, amelyeket eddig emberek végeztek el, egyre inkább NLP-alapú robotok fogják ellátni. Azok a KJI-k, amelyek nem használják ki ezeket az átütő hatású technológiákat, előbb-utóbb *elveszíthetik ügyfeleiket azok javára, amelyek képesek ilyen megoldásokkal hatékonyabban működni*.

⁴⁹ Jian Jia – Ginger Zhe Jin – Liad Wagman: *The Short-Run Effects of GDPR on Technology Venture Investment*. Working Paper 25248. National Bureau of Economic Research, 2018.

Befektetés, tőkekoncentráció

A hatékony NLP-alkalmazások fejlesztéséhez, fenntartásához és támogatásához magas szintű tőkekoncentrációra van szükség, amelyet a KJI-k nem feltétlenül engedhetnek meg maguknak. Minél fejlettebb a piac, annál nagyobb a tőke koncentrációja. Ezért a KJI-knek ennek ellensúlyozására csak egy lehetőségük van, ha időszakosan számlázott *felhőalapú megoldásokra támaszkodnak* ezek minden előnyével és kockázatával együtt. A KJI-knek ugyanazt a felhőalapú alkalmazáscsomagot használva ugyanakkor jelentős erőfeszítéseket kell tenniük, hogy *megkülönböztessék magukat a piacon*.

Jogi felelősség

Az NLP-alkalmazások használatából fakadó *jogi felelősség* megvitatása túlfeszítené a jelen írás kereteit, de jeleznünk kell, hogy ez is jelenthet akadályt. Analógiaként ugyanakkor rámutatunk az önvezető autókkal kapcsolatos jogi problémákra.

Az emberi tényezőből fakadó akadályok

A szakmák jövője az átütő hatású technológiák fényében

A szakmák jövőjéről a mesterséges intelligencia fényében roppant különböző vélemények jelentek meg. Ezekből érdemes akár indirekt következtetéseket is levonni. A *McKinsey* tanácsadó cég azt jósolja, hogy 2030-ra a legrosszabb forgatókönyv szerint akár 800 millió munkahely is megszűnhet, és 375 millió embernek kell más szakmát, hivatást keresnie.⁵⁰ Az *MIT Technology Review* előrejelzése szerint a fehér-galléros munkahelyek egytizedét a robotok fogják elfoglalni a nem túl távoli jövőben.⁵¹ A *Gartner* 2019-ben azt jósolta, hogy 2023-ig a szervezeten belüli kérdések negyedét jogi robotok vagy virtuális jogi asszisztensek (*virtual legal assistant*, VLA) fogják kezelni, helyettesítve azokat a humán jogi asszisztenseket, akik addig elvégezték ezt a munkát.⁵²

Ugyanakkor emlékezni szeretnénk a lóvontatású buszokkal történő londoni közlekedés jóslatára. A lineáris extrapoláció azt jelezte, hogy ha a városi közlekedés volumene a becslés készítésekor tapasztalt sebességgel nő, akkor rövid idő alatt a lóü-

⁵⁰ McKinsey Global Institute: Jobs Lost – Jobs Gained: Workforce Transitions in a Time of Automation. 2017. december.

⁵¹ Will Knight: Is Technology About to Decimate White-Collar Work? *MIT Technology Review*, 2017. november 6.

⁵² Gartner Predicts That by 2023 „Lawbots” Will Handle a Quarter of Internal Legal Requests. *Gartner*, 2019. január 30.

rülék ellepi a várost,⁵³ ami nyilván nem történt meg a belső égésű motorok terjedése okán. Ezzel nem akarjuk a McKinsey-tanulmányt azzal vádolni, hogy szimpla lineáris extrapolációt alkalmazott volna, mindössze célunk egy derűs epizódra a jövő kutatás történetéből.

Optimistább guruk nem a munkahelyek számának jelentős változását, hanem a munkaerőpiac szerkezetének jelentős *átstrukturálódását* hangsúlyozzák: a robotok egyre több rutinmunkát fognak elvégezni, míg az emberek több időt kapnak a kifinomultabb, kreatív feladatokhoz. A jelen írás szerzői, elismerve McKinsey kompetenciáját, kevésbé disztópikus nézetet vallanak. Az 1. táblázat azt mutatja, hogy a leginkább iparosodott országok, ahol a gyártás nagymértékben automatizált, alacsony munkanélküliségi rátával rendelkeznek. Önmagában a képzett munkaerő korlátozott rendelkezésre állása is erős motivációt jelenthet az automatizálás számára. A legkevésbé fejlett országok automatizálási szintjéről nem volt megbízható adat, azonban magabiztosan feltételezhető, hogy ez sehol sem áll közel a felső csoportban lévőkéhez. Az adatokból nem állapítható meg egyértelmű lineáris inverz arányosság az automatizálás és a munkanélküliség között; a tendencia azonban nyilvánvaló: *az automatizálás magas szintje nem eredményez magas munkanélküliséget.*

1. táblázat: A világ tíz leginkább automatizált és a legnagyobb munkanélküliségi rátával rendelkező országa

	Automatizáltság	Munkanélküliség
Dél-Korea	631	4,2
Szingapúr	488	3,6
Németország	309	4,4
Japán	303	2,9
Svédország	223	8,9
Dánia	211	5,7
Egyesült Államok	189	6,2
Olaszország	185	10,7
Belgium	184	5,3
Tajvan	177	3,8
Burkina Faso	n. a.	77
Szíría	n. a.	50
Szenegál	n. a.	48
Haiti	n. a.	40,6
Kenya	n. a.	40
Dzsibuti	n. a.	40
Marshall-szigetek	n. a.	36
Namíbia	n. a.	34
Kiribati	n. a.	30,6

* *Megjegyzések:* A tízezer alkalmazottra jutó telepített robotok száma a gyáripárban; n. a.: nincs adat.
Forrás: a szerzők, Steve Crowe: 10 Most Automated Countries in the World. *The RobotReport*, 2018. február 7. és List of Countries by Unemployment Rate, Wikipedia alapján

⁵³ Ben Johnson: The Great Horse Manure Crisis of 1894. *Historic UK*.

Fontos kérdés, hogy a robotok elterjedése okozhatja-e a jogi szakmában a munkahelyek tömeges megszűnését. Mivel a mesterséges intelligencia térnyerése a jogi világban messze elmarad például a radiológiától, amely a mesterséges intelligencia korai alkalmazója az orvosi területen (nem NLP, hanem képfeldolgozás *Recurrent Neural Network* – RNN és *Decision Tree* osztályozó technológiák révén), érdemes megvizsgálni az analógiát. Az AI-alapú tumorérzékelési teljesítmény jóval meghaladja az emberét. Ennek ellenére a radiológusok nem veszítik el állásukat, csak magasabb képzettséget igénylő feladatokra tudnak összpontosítani. Az 1. táblázat a gépgyártás foglalkoztatási trendjeit tükrözi, és nem szándékozik azt sugallani, hogy számos területen és szakmában nem lesznek fájdalmas strukturális változások, mint például a jogi területen.

A szakképzett munkaerő hiánya

Bizonyossággal állítható, hogy az IKT-képességek szintje a KJI-k esetében a felzárkózáshoz nem elegendő. Nincs kellő számú képzett szakember. A nagy jogi irodák és a nagyvállalatok jogi osztályai a szakemberekért versenyeznek. A KJI-k aligha engedhetik meg maguknak az ilyen drága specialisták alkalmazását, akik egyébként szükségesek lennének a *legalttech* modernizálásához.⁵⁴ Ráadásul *a jogi informatikai szakképzés* még az angolszász egyetemeken is *ritkaság* (illetve most indulnak ezek a képzések), nemhogy az európaiakon. *A magyar egyetemeken pedig ebben a formában nem is létezik.* A jogi informatika tanterve gyakran nem terjed túl az adatbázis-keresési módszertanon. Gyakran a jogi informatikát összekeverik az informatikai joggal.

Rövid távú kontra hosszú távú prioritás

Minden ismert felmérés hangsúlyozza a jogi irodák és jogi osztályok által a jogi informatika fontosságának elismerését. Ugyanakkor a legtöbbjük hangsúlyozza azt a nehézkességet, amely a jogi informatikai alkalmazások bevezetését akadályozza. Az időelszámoláson alapuló munka rövid távú taktikai prioritása gyakran megelőzi a hosszú távú stratégiai beruházásokat.⁵⁵

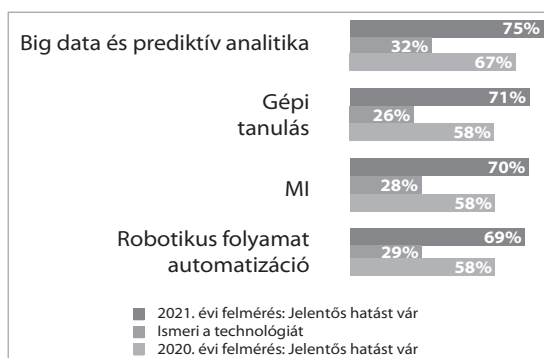
⁵⁴ Frank Ready: 5 Pressing E-Discovery Tech and Personnel Challenges Facing Legal Departments. [Law.com](http://www.law.com), 2021. június 24.

⁵⁵ Legatics: Barriers to Legal Technology Adoption.

A tudás és a tapasztalat hiánya

A technológia elterjedését gátló tényezőként azonosítható továbbá, hogy a vezető beosztású, olykor idősebb személyzet nem rendelkezik elegendő ismeretekkel a jogi informatikai, különösen az NLP-alapú alkalmazásokról. Az ismeretlentől való félelem és a technológiai hozzáértés hiánya gyakran gátolja az innovációt. Ez a félelem és a tétovázás párhuzamba állítható az úgynevezett vezetői vakfoltokkal, ami szintén előfordul a KJI-knél. „Elhiszem, amit el akarok hinni, és figyelmen kívül hagyom, amit nem akarok elhinni.”

A 3. ábra mutatja, hogy a jogi irodák vezetői mennyire tartják fontosnak a jogi informatika használatát, illetve hogy mennyire érzik felkészültnek a szervezetüket a változásra. Érdekes figyelni a változást 2020-ról 2021-re.



3. ábra: A transzformációs technológia becsült hatása és tényleges ismerete

Forrás: *The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report*

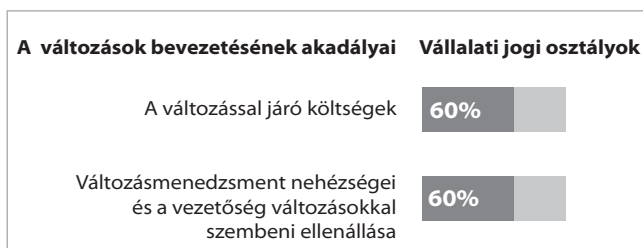
Bizalomhiány, ellenállás a változással szemben

Félelem az ismeretlentől, a megfelelő képzés és tudás hiánya, a sietve bevezetett, rosszul illeszkedő, nem kellőképpen kiértett szoftvertermékek, valamint a bonyolult és nehezen navigálható felhasználói felületek, a túl sok hamis pozitív és – a jogi gyakorlatban rendkívül veszélyes – hamis negatív eredmények bizalmatlansághoz és ezáltal halasztgatáshoz vezetnek.

A mesterséges intelligencia működésének nem megfelelő ismerete miatt a hamis elvárások elégedetlenséget szülnek. A mesterséges intelligencia nem egyszerű szövegszerkesztő programként működik, az alkalmazásnak szüksége van a tanítóadatok folyamatos frissítésére, a finomhangolásra és a visszajelzésre. A karbantartás hiánya szintén az eredmények romlását és így elégedetlenséget okozhat.⁵⁶

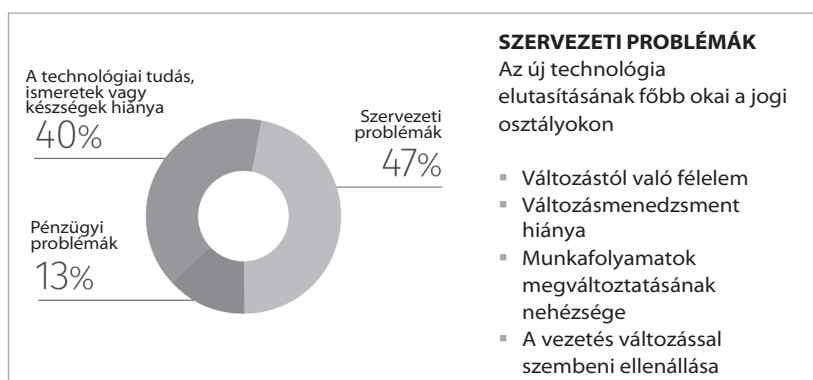
⁵⁶ Commission on the Future of Legal Services: *Report on the Future of Legal Services in the United States*. American Bar Association, 2016.

A Wolters Kluwer 2020-as felmérése szerint a változások kezelésének nehézségei és a vezetésnek a változással szembeni ellenállása jelenti a legnagyobb akadályt az ügyvédi irodák és a vállalati jogi osztályok számára a változások bevezetésében.⁵⁷ Bár nem találtunk kifejezetten KJI-kre vonatkozó adatokat, a WK mérései még mindig egyértelműen jelzik a rezisztenciátényezőket a jogi osztályokon (4–5. ábra), valamint a jogi irodákban (6. ábra).



4. ábra: A változás bevezetésének akadályai

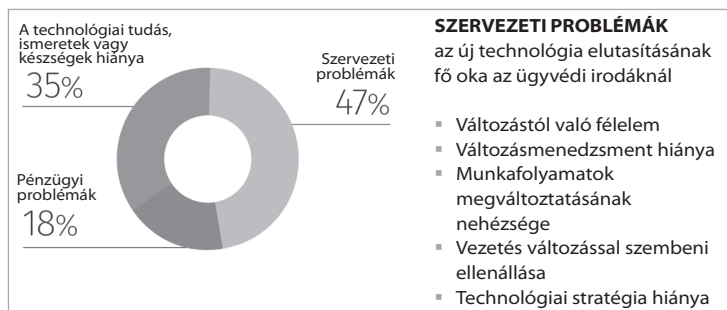
Forrás: *The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report*



5. ábra: Az új technológia jogi irodákban való ellenállásának okai

Forrás: *The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report*

⁵⁷ 2020 Wolters Kluwer Future Ready Lawyer: Performance Drivers and Change in the Legal Sector – Technology Is at the Centre of Top Performance. *Wolters Kluwer*, 2020. július 16.



6. ábra: Az új technológia jogi irodákban való ellenállásának okai

Forrás: The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report

Etikai megfontolások, algoritmikus torzítások

Az AI és ezen belül az NLP működése nem lehet teljesen torzulásmentes. Ennek fő oka a hibás, aránytalanul összeállított tanítóadatokból előállított modell. A mesterséges intelligenciának ugyanúgy kerülnie kellene az etnikai, nemi, szexuális irányultságon, nemi identitáson, fajon, politikai irányultságon, pénzügyi helyzeten stb. alapuló megkülönböztetést, ahogy a humán döntéshozóknak is. A torzítások csökkentésének számos módja van, a teljes kiküszöbölése azonban elméletileg is lehetetlen. Karen Hao elemzi a COMPAS egyesült államokbeli, visszaesési kockázatok bírósági előrejelzését végző szoftverének esetét, és lenyűgöző paradoxont mutat be a témában.⁵⁸

Lehetőségek

Technikai lehetőségek

Dedikált NLP-eszközök KJI-khoz

A digitális átállás kezelésekor érdemes megfontolni a KJI-k számára speciális NLP-alkalmazások előállítását. Ezeknek felhasználóbarátnak, könnyen megtanulhatónak és olcsónak kell lenniük. Ezeknek az alkalmazásoknak – amelyek leginkább felhőalapú szolgáltatásként (*software as a service* – SAAS) működnek – minden olyan területet le kell fedniük, amelyre a KJI-knek a mindennapi üzletmenetük során szükségük lehet. Bár nem KJI-specifikus, a gyorsan növekvő érdeklődés látható a 8–9. ábrán.

⁵⁸ Karen Hao – Jonathan Stray: Can you Make AI Fairer than a Judge? Play Our Courtroom AI-Algorithm Game. *MIT Technology Review*, 2019. október 17.



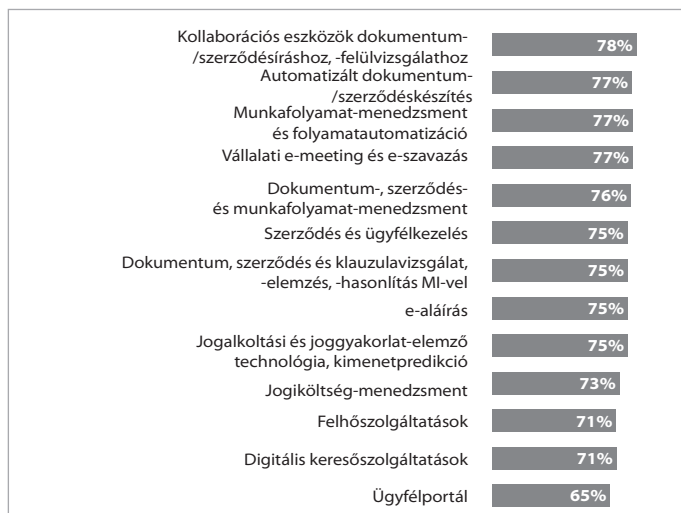
7. ábra: A vállalati jogi osztályok tervezett beruházási irányai

Forrás: The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report

Az adat- és kommunikációbiztonság kiemelten fontos a tárolt ügyfél-információk okán.⁵⁹ A hackerek támadásai, a belső forrás szivárogtatása vagy akár kormányzati beavatkozás miatt egy ügy akár össze is omlhat. A nagyobb jogi irodák megengedhetik maguknak egy teljes belső hálózat, intranet működtetését megfelelő biztonsági intézkedésekkel és eszköztárral. Azonban még a nagy jogi irodák vagy a vállalati jogi osztályok sem támaszkodhatnak teljes mértékben a saját hálózaton belül futtatott alkalmazásokra, és nem zárkoznak el a felhőalapú szolgáltatásoktól. Ebben a dokumentumban nincs lehetőség a titkosítási módszerek és termékek ismertetésére. Meg kell azonban jegyezni, hogy a kvantumszámítógépekkel a jövőben könnyen fel lehet majd törni a jelenleg biztonságosnak tartott titkosításokat is. Ez az átütő hatású technológia sok tekintetben meg fogja változtatni az IKT jelenlegi működését.

⁵⁹ Leena Iyar: Small Legal Firms Are Struggling with the Adoption of Digital Platforms. *ABA*, 2021. április 27.

A természetes nyelvi feldolgozóeszközök kis jogi irodákban történő használatának akadályairól és lehetőségeiről



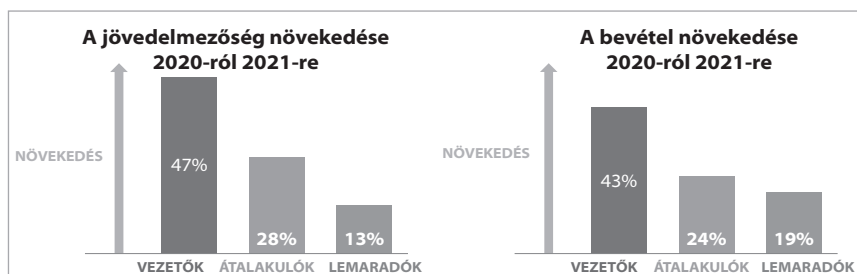
8. ábra: Technológiafejlesztési prioritások

Forrás: The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report

Emberi és szervezeti lehetőségek

A felhasználható alkalmazások körét korlátozza a KJI-k pénzügyi teherbírása, az alkalmazások/szolgáltatások és a szakértelem rendelkezésre állása, valamint a piaci nyomás, beleértve az ügyfelek preferenciáit is.

A Wolters Kluwer 2020-as felmérése szerint a válaszadó ügyvédi irodák, a vállalati jogi osztályok és az üzleti szolgáltató cégek 76%-a Európában és az Egyesült Államokban a *legaltechet* tartja a legmeghatározóbb faktornak. Ugyanakkor mindössze csak egyharmaduk érezte magát felkészültnek a változásokra.⁶⁰



9. ábra: A technológiai vezetők jobban teljesítenek

Forrás: The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report

⁶⁰ 2020 Wolters Kluwer Future Ready Lawyer: Performance Drivers and Change in the Legal Sector – Technology Is at the Centre of Top Performance. Wolters Kluwer, 2020. július 16.

A WK felmérése (9. ábra) azt mutatja, hogy a jogi technológiában befektetni hajlandó irodák jövedelmezősége és növekedése messze meghaladja a lemaradókét.

Ajánlások és összefoglalás

Mint fentebb említettük, az NLP-megoldásokhoz funkcionális nyelvi modellre van szükség, amely viszont elegendő képzési adatot igényel. Ezért még a kis nyelveknek (beleértve a kis többközpontú nyelveket) is elegendő jogi korpuszra van szükségük ahhoz, hogy táplálják nyelvi modelljüket, ami egyszerűen nem érhető el. Ezen nyelvek digitális jelenlétének megteremtése és megőrzése érdekében támogatni kell ezeket az erőfeszítéseket. A nyelvi modellek, különösen a kis nyelvek legális nyelvi modelljei feltétlenül szükségesek ahhoz, hogy a kis népesség ne essen az út szélére.

Nagyon ajánlott, hogy egy részletes felmérés feltárja a nyelvi modellek helyzetét általában, és konkrétan a jogi modellekét, országonként az EU-ban, különös tekintettel a kis nyelvekre (beleértve a kis többközpontú nyelveket). Az eredmények alapján javaslatot kell benyújtani az Európai Kutatási Tanácsnak e számítástechnikai nyelvi tevékenységek támogatása céljából.

Programokat kell létrehozni a fejlettebb országok számára, hogy másokat támogassanak a szoftver- és nyelvi modellek fejlesztésében. Ennél is fontosabb a technológiaátadási kompetencia-központok regionális létrehozása a kisegítő személyzet képzése a velük folytatott konzultáció és a munkájuk támogatása céljából. Be kell vezetni a jogi szakképzést, vagy annak szintjét jelentősen meg kell emelni az EU valamennyi egyetemén. Az egyetemek közötti együttműködést minden ismert módon támogatni kell.

Figyelmeztetésként érdemes megemlíteni, hogy a védekező stratégiák protekcionista lépésként fejleszthetők ki és valósíthatók meg a szabályozási környezetben. Ezek egyelőre munkahelyeket és pozíciókat takaríthatnak meg, hosszú távon azonban csak a technológiailag megőrzött status quót fogják fenntartani.

Kérdésként vetődik fel (jelen kötet *Bevezetése* is az egyik kulcskérdésként tárgyalja), hogy az AI helyettesítheti-e, kiválthatja-e az ügyvédek. Számos tanulmány foglalkozik a jogi szakma jövőjével. Egyértelműen hiány van NLP-megoldásokra specializálódott, magasan képzett jogi szakértőkből. Az úgynevezett *General AI* nagyon távol áll attól, hogy olyan jogi szakértőket helyettesítsen, akik képesek gondolkodni, szintetizálni, vezetni és kommunikálni. Az ügyvédi szakma egészét nem fenyegeti tehát közvetlen veszély, de bizonyos tevékenységeket az AI-alkalmazások már rövid távon is el fognak tudni végezni.

A jogi szakirodalom alig foglalkozik az átütő hatású technológiák használata terén a kis ügyvédi irodákra jellemző akadályokkal és lehetőségekkel. A legtöbb tanulmány a nagy ügyvédi irodákra és jogi osztályokra összpontosít, vagy nem tesz különbséget az ügyvédi iroda mérete szerint.

A fő akadályok nem technológiai, hanem emberi és szervezeti jellegűek.

A kis nyelvek veszélyeztetettek. Elkötelezett erőfeszítések nélkül a digitális jelenlétük még a legális iparban is, csökkenhet.

A költségmegtakarítások miatt az ügyfelek érdeklődése az ügyvédi irodák felől az alternatív jogi szolgáltatók felé tolódik el. A kiszervezett tevékenységek nem igényelnek magas szintű képzettséget, de könnyen automatizálhatók.

A jogi mesterséges intelligenciát a legtöbb forrás polarizált módon szemléli: sok szempontból túlbecsülik vagy éppen alábecsülik.

A KJI-ket általában nem veszélyezteti az NLP; azok azonban, amelyek nem képesek befogadni az új technológiákat, kihívások elé fognak nézni az elkövetkező években.

A KJI-k oktatása kulcsfontosságú. Képzési és IKT-támogató központokra van szükség. Az NLP és más korszerű technológiák oktatása a jogi egyetemeken elengedhetetlen.

Felhasznált irodalom

- 2020 Wolters Kluwer Future Ready Lawyer: Performance Drivers and Change in the Legal Sector – Technology Is at the Centre of Top Performance. *Wolters Kluwer*, 2020. július 16. Online: www.wolterskluwer.com/en/expert-insights/performance-drivers-and-change-in-the-legal-sector
- Brush, Kate – Jesse Scardina: Chatbot. Definition. *TechTarget*. Online: www.techtarget.com/searchcustomerexperience/definition/chatbot
- Commission on the Future of Legal Services: *Report on the Future of Legal Services in the United States*. American Bar Association, 2016. Online: www.srln.org/system/files/attachments/2016%20ABA%20Future%20of%20Legal%20Services%20Report-Web.pdf
- Crowe, Steve: 10 Most Automated Countries in the World. *The RobotReport*, 2018. február 7. Online: www.therobotreport.com/10-automated-countries-in-the-world/
- Dale, Robert – Jette Viethen: The Automated Writing Assistance Landscape in 2021. *Natural Language Engineering*, 27. (2021), 4. 511–518. Online: <https://doi.org/10.1017/S1351324921000164>
- Das, Sejuti: GPT-3 Vs BERT For NLP Tasks. *Analytics India Magazin*, 2020. szeptember 11. Online: <https://analyticsindiamag.com/gpt-3-vs-bert-for-nlp-tasks/>
- Does Grammarly Support Languages Other than English? *Grammarly Support*. Online: <https://support.grammarly.com/hc/en-us/articles/115000090971-Does-Grammarly-support-languages-other-than-English>
- Dokumentum-osztályozás. *OpinHU Tartalomelemzés*. Online: <https://sites.google.com/a/geox.hu/opinhu/hatteranyagok/fogalomtar/dokumentum-osztalyozas>
- Entitásfelismerés. *OpinHU Tartalomelemzés*. Online: <https://sites.google.com/a/geox.hu/opinhu/hatteranyagok/fogalomtar/entitasfelismeres>
- European Commission: *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. COM(2020)65 végleges (2020. február 19.). Online: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Executive Summary der Rechtsdienstleistungsmarkt 2030*. Deutscher Anwalt Verein – Prognos. Online: <https://anwaltverein.de/de/anwaltspraxis/dav-zukunftsstudie?file=files/anwaltverein.de/downloads/service/DAV-Zukunftsstudie/DAV-Zukunftsstudie-Executive-Summary.pdf>

- Gartner Predicts That by 2023 “Lawbots” Will Handle a Quarter of Internal Legal Requests. *Gartner*, 2019. január 30. Online: www.gartner.com/en/newsroom/press-releases/2020-01-30-gartner-predicts-that-by-2023-lawbots-will-handle-a-quarter-of-internal-legal-request
- Gépi fordítás. *Wikipédia*. Online: https://hu.wikipedia.org/wiki/Gépi_fordítás
- Ghavidel, Hadi – Amal Zouaq – Michel Desmarais: Using BERT and XLNET for the Automatic Short Answer Grading Task. In *Proceedings of the 12th International Conference on Computer Supported Education*. I. CSEDU, 2020. 58–67. Online: <https://doi.org/10.5220/0009422400580067>
- Glaser, Ingo – Sebastian Moser – Florian Matthes: Sentence Boundary Detection in German Legal Documents. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence*. II. ICAART, 2021. 812–821. Online: <https://doi.org/10.5220/0010246308120821>
- Hao, Karen – Jonathan Stray: Can you Make AI Fairer than a Judge? Play Our Courtroom Algorithm Game. *MIT Technology Review*, 2019. október 17. Online: www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm
- Henderson, William D.: Innovation Diffusion in the Legal Industry. *Dickinson Law Review*, 122. (2018). 2. 395–478.
- How Many Languages Are There in the World? *Ethnologue*. Online: www.ethnologue.com/guides/how-many-languages
- Iyar, Leena: Small Legal Firms Are Struggling with the Adoption of Digital Platforms. *ABA*, 2021. április 27. Online: www.americanbar.org/groups/business_law/publications/blt/2021/05/digital-platforms/
- Jia, Jian – Ginger Zhe Jin – Liad Wagman: *The Short-Run Effects of GDPR on Technology Venture Investment*. Working Paper 25248. National Bureau of Economic Research, 2018. Online: <https://doi.org/10.3386/w25248>
- Johnson, Ben: The Great Horse Manure Crisis of 1894. *Historic UK*. Online: www.historic-uk.com/HistoryUK/HistoryofBritain/Great-Horse-Manure-Crisis-of-1894/
- Knight, Will: Is Technology About to Decimate White-Collar Work? *MIT Technology Review*, 2017. november 6. Online: www.technologyreview.com/2017/11/06/147955/is-technology-about-to-decimate-white-collar-work/
- Kornai, András: Digital Language Death. *PLOS ONE*, 2013. október 22. Online: <https://doi.org/10.1371/journal.pone.0077056>
- Languages of Europe. *Wikipedia*. Online: https://en.wikipedia.org/w/index.php?title=Languages_of_Europe&oldid=1036009595
- Legatics: *Barriers to Legal Technology Adoption*. Online: <https://events.legatics.com/hubfs/Barriers%20to%20Legal%20Technology%20Adoption.pdf>
- List of Countries by Unemployment Rate. *Wikipedia*. Online: https://en.wikipedia.org/wiki/List_of_countries_by_unemployment_rate
- List of Endangered Languages in Europe. *Wikipedia*. Online: https://en.wikipedia.org/w/index.php?title=List_of_endangered_languages_in_Europe&oldid=1035831048
- List of Wikipedias. *Wikimedia*. Online: https://meta.wikimedia.org/wiki/List_of_Wikipedias
- Markhardt, Heidemarie: *Das Österreichische Deutsch im Rahmen der EU*. Frankfurt am Main – Berlin – Bern – Bruxelles – New York – Oxford – Wien, Peter Lang, 2005.
- Matthews, Kayla: Applications for Legal Chatbots. *Law Technology Today*, 2020. július 13. Online: www.lawtechnologytoday.org/2020/07/applications-for-legal-chatbots/

A természetes nyelvi feldolgozóeszközök kis jogi irodákban történő használatának akadályairól és lehetőségeiről

- McKinsey Global Institute: *Jobs Lost – Jobs Gained: Workforce Transitions in a Time of Automation*. 2017. december. Online: www.mckinsey.com/~media/BAB489A30B724BECB5DEDC41E9BB9FAC.ashx
- Muhr, Rudolf: Die plurizentrischen Sprachen Europas – ein Überblick. In Eva Gugenberger – Mechthild Blumberg (szerk.): *Europa. Zur Situation der regionalen Sprachen von der Iberischen Halbinsel bis zum Kaukasus*. Österreichisches Deutsch – Sprache der Gegenwart, II. Frankfurt am Main, Peter Lang, 2003. 191–233.
- Muhr, Rudolf – Josep-Àngel Mas Castells – Jack M. Rueter: *European Pluricentric Languages in Contact and Conflict*. New York, Peter Lang, 2020. Online: <https://doi.org/10.3726/b16182>
- Muhr, Rudolf: The Differences in the Legal Terminology of Austria and Germany and the Results for the German Legal Terminology within the Scope of the European Union. *Muttersprache*, 119. (2009), 3. 199–216.
- Multilingual Statistics. *Wikipedia*. Online: https://en.wikipedia.org/w/index.php?title=Wikipedia:Multilingual_statistics&oldid=944342501
- Netjes, Willemien – Arno R. Lodder: e-Court – Dutch Alternative Online Resolution of Debt Collection Claims. *International Journal of Online Dispute Resolution*, (2019), 1. 96. Online: <https://doi.org/10.5553/IJODR/235250022019006001005>
- Panetta, Kasey: 5 Trends Appear on the Gartner Hype Cycle for Emerging Technologies, 2019. *Gartner*, 2019. augusztus 29. Online: <http://gtmr.it/3uuYs4J>
- Park, Andrew Lee: Injustice Ex Machina: Predictive Algorithms in Criminal Sentencing. *UCLA Law Review*, 2019. február 19. Online: www.uclalawreview.org/injustice-ex-machina-predictive-algorithms-in-criminal-sentencing/
- Pluricentric Language. *Wikipedia*. Online: https://en.wikipedia.org/wiki/Pluricentric_language
- Prakash Maheswari, Jyoti: Breaking the Curse of Small Data Sets in Machine Learning: Part 2. *Medium*, 2019. április 16. Online: <https://towardsdatascience.com/breaking-the-curse-of-small-data-sets-in-machine-learning-part-2-894aa45277f4>
- Ready, Frank: 5 Pressing E-Discovery Tech and Personnel Challenges Facing Legal Departments. *Law.com*, 2021. június 24. Online: www.law.com/legaltechnews/2021/06/24/5-pressing-e-discovery-tech-and-personnel-challenges-facing-legal-departments/?slreturn=20220030083513
- Ready, Frank: Legal Departments Likely Trust ALSPs More than Firms on Tech Recommendations. *Law.com*, 2021. augusztus 4. Online: www.law.com/legaltechnews/2021/08/04/legal-departments-likely-trust-alsps-more-than-firms-on-tech-recommendations/?kw=Legal%20Departments%20Likely%20Trust%20ALSPs%20More%20Than%20Firms%20on%20Tech%20Recommendations
- Riekert, Martin – Matthias Riekert – Achim Klein: Simple Baseline Machine Learning Text Classifiers for Small Datasets. *SN Computer Science*, 2. (2021). 178. Online: <https://doi.org/10.1007/s42979-021-00480-4>
- Sanchez, George: Sentence Boundary Detection in Legal Text. In *Proceedings of the Natural Language Processing Workshop 2019*. Association for Computational Linguistics, 2019. 31–38. Online: <https://doi.org/10.18653/v1/W19-2204>
- Sentiment Analysis. *Wikipedia*. Online: https://en.wikipedia.org/wiki/Sentiment_analysis
- Size and Vitality of Shan. *Ethnologue*. Online: www.ethnologue.com/size-and-vitality/shn
- The 2021 Wolters Kluwer Future Ready Lawyer. Moving Beyond the Pandemic. Survey Report*. 2021. Online: https://images.go.wolterskluwer.com/Web/WoltersKluwer/%7B53280717-ca9f-47f5-bff6-a35e52230ef5%7D_FRL2021_WhitePaper_Global_FINAL.pdf

- The Bellwether Report Series 2019: Is the Future Small? *LexisNexis*. Online: www.lexisnexis.co.uk/research-and-reports/is-the-future-small-bd.html
- Turing-teszt. *Wikipédia*. Online: <https://hu.wikipedia.org/wiki/Turing-teszt>
- Vadász János Pál – György Görög – Csaba Csáki – István Üveges: *Opportunities and Barriers in the Use of Natural Language Processing Tools in the SME Law Practices*. JUST-JACC-EJU-AG-2019, grant 881527. Online: <https://elf-fae.eu/wp-content/uploads/2021/12/Report-on-opportunities-and-barriers-in-the-use-of-NLP-tools-in-SME-law-practices.pdf>
- Vadász, Pál – András Benczúr – Géza Füzesi – Sándor Munk: Identification Illegal Cartel Activities from Open Sources. In Babak Akhgar – Petra Saskia Bayerl – Fraser Sampson (szerk.): *Open Source Intelligence Investigation*. Cham, Springer, 2016. 251–273. Online: https://doi.org/10.1007/978-3-319-47671-1_16
- Vincze Veronika: A Miskolc Jogi Korpusz nyelvi jellemzői. In Szabó Miklós – Vinnai Edina (szerk.): *A törvény szavai*. Miskolc, Bíbor Kiadó, 2018. 9–36.
- Wenjie, Zhang: *The Comparison between the Tools for Named Entity Recognition*. Master's Thesis. Auckland University of Technology, 2020.
- What is an Alternative Legal Service Provider (ALSP)? *ParalegalEDU.org*. Online: www.paralegaledu.org/alternative-legal-service-providers/
- Zahorsky, Rachel – William D. Henderson: Who's Eating Law Firms' Lunch? *ABA Journal*, 2013. október 1. Online: www.abajournal.com/magazine/article/whos_eating_law_firms_lunch
- Zódi, Zsolt: Algorithmic Explainability and Legal Reasoning. *The Theory and Practice of Legislation*, 10. (2022), 1. 67–92. Online: <https://doi.org/10.1080/20508840.2022.2033945>