

Zódi Zsolt¹

A big data alapú jogi elemzőprogramok



Absztrakt: A big data alapú jogi elemzőprogramok (*legal analytics*) a 2010-es évek elején jelentek meg az USA-ban. Úgy működnek, hogy általában weboldalakon található nyilvános adatbázisokból nagy mennyiségű szöveges adatot gyűjtenek össze, ezeket az adatokat tisztítják, gazdagítják, más adatokkal összekapcsolják, majd speciális szoftverekkel elemzik és prezentálják.

Ebben a fejezetben előbb bemutatom az elemzőprogramok történetét, különböző válfajait és a legfontosabb szolgáltatásokat, és ezek tartalmi és technológiai alapjait. Egy kiemelt példán, a „bíróelemző” modulon keresztül a felhasználási lehetőségeket is ismertetem. Még ebben a részben írok arról, hogy a bíróelemző szolgáltatás milyen port kavart Franciaországban, milyen érvek és ellenérvek hangzottak el az e körül zajló vitában.

A második részben a tanulmány az elemzőprogramok jogban történő felhasználási lehetőségeit és az ezzel kapcsolatos aggályokat ismerteti. Az aggályokat a tanulmány két csoportba sorolja. Az első típusú – általános – aggály az „adatalapú jog” és a „predikció” jogban történő korlátozott használatának lehetőségére mutat rá. A második típusú ellenérv pedig a *common law* és a kontinentális jogrendszerek különbségét helyezi a középpontba. A tanulmány végül röviden az elemzők jövőjét elemzi.

Tárgyszavak: big data a jogban, big data a jogalkalmazásban, jogi elemzőprogramok, adatalapú jog, bíróelemző, bíróelemző betöltése Franciországbán, predikció a jogban, Emberi Jogok Európai Bírósága, szövegbányászat a jogban, RegEx, Named Entity Recognition, webscraping, bag of words (BoW), korpusznyelvészet, jog mint adat

A big data alapú jogi elemzőprogramok

A big data kifejezés a 2010-es évek elején terjedt el.² Azt a folyamatot jelöljük vele, amelynek során a főként a nyílt internetről és a dolgok internetéről hozzáférhető egyre több adat segítségével az élet különböző területein új típusú elemzésekre, összefüggések felismerésére, következtetések levonására és előrejelzések készítésére nyílik lehetőség.

¹ Zódi Zsolt életrajzát lásd a kötet Bevezetésének első lábjegyzetében.

² Viktor Mayer-Schönberger – Kenneth Cukier: *Big Data. The Essential Guide to Work, Life, Learning in the Age of Insight*. London, John Murray, 2013.

A big data jelenséget elsősorban szemléletmódként és narratívaként³ lehet a legjobban megragadni, amely a régebb óta ismert statisztikai módszereket igyekszik alkalmazni olyan új területeken, ahol eddig nem állt rendelkezésre kellő mennyiségű adat, így például az üzleti életben a vevők viselkedésének előrejelzésére vagy bizonyos társadalmi jelenségek bekövetkeztének megjósolására. A big data a mesterséges intelligencia alkalmazásában is fordulatot hozott, hatásai közül talán ez a legnagyobb. Az élet különböző területein gyűjtött nagy mennyiségű adat ugyanis tanítóadatként is funkcionál gépi tanuláson alapuló alkalmazásokhoz. Ez sok területen (például a gépi fordítás vagy a képalkotó eljárásokon alapuló orvosi diagnosztika területén) igazi áttöréseket hozott a 2010-es évek második felében.

A big data szemléletmód igen hamar megjelent a jog világában is.⁴ Sokan méltatták a benne rejlő lehetőségeket, amelyek segíthetik a jogalkotást (például valós idejű adatokra és azonnali visszacsatolásra alapozott jogalkotási „kísérletezéssel”), a jogtudományt (például bírósági ítéletek szövegbányászati elemzésével) és a jogalkalmazást is, akár bírósági, akár hatósági jogalkalmazásról van szó.⁵ Az itt következő írás ez utóbbról szól: az elemzőprogramok alkalmazásáról és ennek korlátairól a jogalkalmazás területén.

Az elemzőprogramok története

Amerikában már a 2000-es évek közepén elindultak azok a kísérleti projektek, amelyekben a hozzáférhető adatok újszerű összesítésével és feldolgozásával új típusú eszközt igyekeztek a jogalkalmazók, elsősorban az ügyvédek kezébe adni.

Az első ilyen elemzőprogram a Lex Machina volt, amelynek közvetlen előzménye a Stanford Law School és a Stanford Computer Science Department együttműködésében jött létre, még 2006-ban. A projekt háttérében egy másik kezdeményezés állt, az IP (Intellectual Property) Litigation Clearinghouse, amely a szellemi tulajdonnal (elsősorban a szabadalmakkal) kapcsolatos jogviták keresőrendszerét szerette volna létrehozni.⁶ A szellemi alkotások joga azért különösen jó terület a kísérletezésre,

³ Lásd például Zödi Zsolt: A jogi szövegek mint big data. In Jakab András – Sebők Miklós (szerk.): *Empirikus jogi tanulmányok*. Budapest, Osiris Kiadó – MTA Társadalomtudományi Kutatóközpont, 2020. 93–109.

⁴ Nancy B. Rapoport – Joseph R. Tiano Jr.: Legal Analytics, Social Science, and Legal Fees: Reimagining Legal Spend Decisions in an Evolving Industry. *Georgia State University Law Review*, 35. (2019), 4. 1269–1304.; Zödi Zsolt: Jog és jogtudomány a big data korában. *Allam- és Jogtudomány*, 58. (2017), 1. 95–114.

⁵ Daniel Martin Katz: Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*, 62. (2013), 4. 909–966.

⁶ Patrick Hoge: Lex Machina: 'Moneyball' Meets Patent Lawsuits. *San Francisco Business Times*, 2013. szeptember 3.

mert a jogterület körül nagy tőkék mozognak, a pereskedés általában nagyon nagy költségekkel jár, és az ügyek egy viszonylag szűk tárgyterületet (domaint) érintenek, így a bennük használt érvkészlet jól lehatárolható. (Itt emlékeztetünk arra, hogy az egyik első sikeres jogi szakértői rendszer, a HYPO az üzletititok-sértés területén működött, amely hasonló jellemzőkkel bírt.⁷) A Lex Machina 2015-ig töretlenül gyűjtötte a befektetőket és az ügyfeleket, hogy 2015-ben aztán felvásárolja a Lexis-Nexis.⁸

Ezzel egy időben más szolgáltatók is feltűntek. Jelenleg a 360quadrants⁹ weboldal szerint 18 cég kínál ilyen szolgáltatásokat, bár úgy tűnik, ezek nem feleltethetők meg egy az egyben egymásnak. Itt fontos megjegyezni, hogy az elemzőkben található funkcionális egy része más típusú szoftverekben is megjelenik. Az „általános” elemzőkön kívül léteznek olyan elemzőprogramok is, amelyek az *e-Discoveryt* vagy a *legal due diligence-et* segítik.¹⁰ Utóbbiak funkcionálisa tehát nagyon hasonló (keresés, aggregálás/vizualizálás, predikció), de ezek nem nyilvános jogi dokumentumokon dolgoznak, hanem az ügyfél saját dokumentumain, illetve az ezekből kinyert adatokat olykor összehasonlítják a nyilvános adatbázisokban található adatokkal.

A jogi elemzők tehát egyértelműen köthetők a nyílt interneten megjelenő nagy tömegű adathoz, és az angolszász jogi kultúrához, ahol a per (sőt minden jogi eljárás, akár egy szerződéskötést megelőző tárgyalás is) kognitív, érvelési csata, amelyben az nyer, akinek több információja van.

Technológia, funkciók

A jogi elemzőszolgáltatások tehát jogi dokumentumokat elemeznek. A dokumentumok, ahogy azt fentebb említettem, lehetnek nyilvános dokumentumok valamilyen nyílt adatbázisból vagy a felhasználó saját dokumentumai.¹¹ Előbbiket a legtöbb szolgáltató ún. *webscraping* technológiákkal – azaz leegyszerűsítve:

⁷ Kevin D. Ashley: Reasoning with Cases and Hypotheticals in HYPO. *International Journal of Man-Machine Studies*, 34. (1991), 6. 753–796.

⁸ LexisNexis Acquires Premier Legal Analytics Provider Lex Machina. *Lex Machina*, 2015. november 23.

⁹ Best Legal Analytics Software 2021.

¹⁰ Az e-Discovery rendszerekről lásd a Bevezetést, illetve Vági Renátó tanulmányát.

¹¹ A Lex Machina például négy nyilvános adatbázist használ az adatok begyűjtésekor. A PACER-t (Public Access to Court Electronic Records), az USA szövetségi fellebbviteli bíróságainak és csődbíróságainak teljes iratanyagát tartalmazó adatbázist, az EDIS-t (Electronic Document Information System), a Nemzetközi Kereskedelmi Bizottság (International Trade Commission) döntéseinek és vizsgálatainak iratanyagát, a US PTO (Patent Office – Szabadalmi Hivatal) adatbázisát, és bizonyos állami bíróságok (State Courts) adatbázisait. A cikk írásakor 16 megye bíróságainak dokumentumai találhatók meg a rendszerben, ami – tekintve, hogy az USA-ban jelenleg 3006 megye van – korántsem reprezentatív.

tömeges letöltésekkel – gyűjti be,¹² de természetesen nem kizárt az egyes adat-szolgáltatók közvetlen adatátadása sem. Ezután a szövegeken többféle technikát alkalmaznak, hogy belőlük adatokat és információkat nyerjenek ki. Ezek közül az információkinyerő (*information extraction*) technológiák közül a legtöbbet alkalmazott a *RegEx* (*Regular expression*) és a *named entity recognition*. Előbbi, a „reguláris kifejezés” lényegében szavak (karaktersorok, *stringek*) szabályokkal leírt halmaza. A rendszer a szabályokban definiált karaktersorokat (például egy jogi dokumentumban a jogszabály-hivatkozásokat) megtalálja, majd leindexeli. Utóbbi pedig felfogható az előbbi részhalmozaként, és az a lényege, hogy a strukturálatlan vagy részben strukturált szövegekből tág értelemben vett neveket, azaz „merev jelölőket” (személynevek, intézménynevek, helységnevek) és numerikus adatokat (dátum, pénzösszeg vagy más naturália mennyisége) ismer fel, emel ki, majd tárol el. A két módszert összeköti tehát, hogy a folyó szövegekből *adatokat* emelnek ki. Az elemzők szolgáltatásai, funkciói ezekre az adatokra épülnek.

Ezek a funkciók alapvetően három csoportba sorolhatók: a kereső, az aggregáló/vizualizáló és a prediktív funkciókra.

Először is minden rendszer rendelkezik „hagyományos” keresőfunkciókkal. A gazdag metaadatkészlet (a sokféle kinyert adat) miatt ezek a rendszerek általában jóval pontosabb, összetettebb keresést tesznek lehetővé, mint a szimpla teljes szövegű vagy néhány egyszerű paraméter alapján kereső rendszerek. További jellegzetességük, hogy eltérően az általános internetes keresőktől, széles körben teszik lehetővé a különböző metaadatok/paraméterek előzetes beállítását (az ún. parametrikus keresést) a kereséskor. Az internetes keresőkben ugyanis a fő felületen általában csak egy keresődoboz található, a kereső pedig különféle egyéb adatokból igyekszik „kitalálni”, hogy a kereső szándéka, a „*user intent*” valójában mi, hiszen ez ugyanazon keresőstring esetén is lehet nagyon különböző. Ebből a szempontból érdekes és tanulságos a Google kereső tesztelőinek kiadott kézikönyv,¹³ amelyből kitűnik, hogy a Google egy-egy keresés esetén mindenekelőtt a keresés helyét és a keresési nyelvet alapul véve szűkíti a lehetséges felhasználói szándékokat, majd a „domináns jelentést” veszi alapul, de a felhasználó személyes adatait és keresési előzményeit is figyelembe veszi. Ha például a felhasználó az „*apple*” szót üti be, a kereső domináns jelentésnek az Apple Inc. IT céget veszi, és csak „gyakori” keresési szándékként kezeli a ’gyümölcs’ jelentést.¹⁴ A szakkeresők azonban másképp működnek, mert ezek kizárólag a keresőstringet veszik figyelembe, a felhasználói szándékot nem képesek bemérni. A felhasználói szándék „kifejezésére” valók a különböző metaadatok. Ebben az értelemben az elemzőprogramok keresői ún. vertikális vagy szakkeresők (*vertical, topical search engine*).

¹² Ryan Mitchell: *Web Scraping with Python*. Sebastopol (CA), O’Reilly Media, 2018.

¹³ General Guidelines 2020. (For Search Quality Rating of Google Search Engine) 2021. október 19. Elérhető a Google Search Help oldalról.

¹⁴ General Guidelines 2020 (2021): i. m. 65.

A rendszerek másik funkciója az aggregáló/statisztikakészítő funkció. Ez a hagyományos keresők találati megjelenítőinél (*result list*) annyival nyújt többet, hogy nem információegységeket (például jogszabályokat vagy bírósági ítéleteket, azaz dokumentumokat) ad vissza találatként egy keresésre, hanem az információegységekben (dokumentumokban, rekordokban) található *adatokat* képes összesíteni és előre definiált formában megjeleníteni. Az ilyen aggregáló funkciók közül is a leglátványosabb – és egyben a legvitatottabb is – az, amely *egy bíróra* vonatkozóan összesíti a rendszerben található adatokat, erről később részletesen szólok. De aggregálást természetesen bármilyen más adatra vonatkozóan is lehet készíteni. Egy bírósági határozatokat tartalmazó adatbázis esetében például aggregálhatók és prezentálhatók egy ügyvédi iroda, egy bíróság, egy ügyfél (cég) adatai – például az ügyvédi iroda sikerességi rátája, az ügyvédi iroda nevében eljáró ügyvédek, az iroda ügyfeleinek listája, az iroda és egy bíró „viszonya” stb. Ugyanígy lehetséges, hogy egy adott bíróságot helyezünk a lekérdezés középpontjába: például egy adott típusú ügy előfordulási gyakoriságát, az ügyek átlagos tartamát, és ezen belül az egyes eljárási szakaszok hosszát (perelőkészítés, elsőfokú, másodfokú eljárás stb.), szakértő kirendelésének gyakoriságát stb. Ugyanígy dokumentumok egy csoportjára vonatkozóan is elvégezhető adatgyűjtés és megjelenítés, például egy meghatározott szűkebb jogterület, mint a lakásbérleti szerződésekre vonatkozó ügyek tekintetében: kik járnak el ezekben az ügyekben, milyen hosszú az eljárás időtartama, jellemzően milyen eredménnyel végződik az eljárás (a bérlő vagy a bérbeadó nyer-e tipikusan) és még sok más is. Lényegében minden kinyerhető adat minden más adattal mutatható együtt – ami önmagában nem túl innovatív, hiszen például a relációs adatbázisok lényege épp az, hogy bármilyen adat bármilyen más adattal összekapcsolva lekérdezhető –, az innovatív ezekben a rendszerekben elsősorban az, hogy kinyerik és adatbázisba rendezik a korábban csak strukturálatlan formában rendelkezésre álló adatokat, így azok lekérdezhetővé válnak. Az adatokat természetesen nemcsak egyszerű listák és táblázatok formájában képes a rendszer prezentálni, hanem különböző módon vizualizálni is tudja ezeket: akár egyszerű diagramként, de például az időtartamok esetén Gantt-diagramként vagy más képi megjelenítésben is.

A predikciós funkció

Az elemzőrendszerek harmadik, egyben legvitatottabb funkciója a *predikciós* funkció, ahol többféle paraméter alapján lehet kiválasztani, hogy milyen ügyre és milyen feltételek mellett kívánunk „jóslást” készíteni. A predikció egyébként sok esetben funkcionálisan nem más, mint dokumentumcsoportosítás (besorolja a rendszer egy-fajta döntéstípusba a dokumentumot), máskor százalékos formába foglalt előrejelzés egy ügy vagy ügytípus kimenetelére vonatkozóan. Hogy a dokumentumcsoportosításon alapuló predikciót el tudjuk képzelni, érdemes felidézni egy híres kísérletet,

amely máig az egyik legidézettebb a témában. Itt Aletras és társai¹⁵ az Emberi Jogok Európai Bíróságának ítéleteit vették alapul, ezen belül is a nagyszámú ítélettel értelmezett 3. (kínzás tilalma), 6. (tisztességes tárgyaláshoz való jog) és 8. (magán- és családi élet tiszteletben tartásához való jog) cikkeket, ahol is 250, 80 és 254 ítélet állt rendelkezésre a tanulmány írásának időpontjában. Ezután letöltötték az összes ítéletet, majd feldarabolták az ítélet alcímei mentén, azaz elkülönítették a szöveg-egységeket egymástól.

Ezután egyenlő számban kiválogattak az egyezmény megsértését kimondó és ezzel ellentétes, jogsértést nem megállapító ítéleteket. Minden ítéletből kiszedték az ítélet végeredményére utaló (megsértette/nem sértette meg) információt, majd RegEx használatával, illetve a stopwordok (hulladékszavak, mint a névelő, kötőszó stb.) kiemelésével „lecsupaszították” a szöveget.

Az így lecsupaszított ítéletek minden részéből (emlékezzünk, az ítéleteket az alcímek szerint bontották szét: például tények, pertörténet) kétféle természetes nyelvfeldolgozási módszerrel nyertek ki információt. Egyrészt az ún. szózsákmódszerrel (*Bag of Words* – BOW) minden grammatikai és szintaktikai sajátosság nélkül kerestek és tároltak *szó-együttelőfordulási gyakoriságokat*. 2000 ún. N-gramot találtak és számoltak meg, mindegyiket külön-külön az egyes ítéletrészekre vonatkozóan. (Az N-gram egy meghatározott szócsoporthoz, kifejezés, szófordulat.) Másrészt ún. témákat (*topic*) képeztek az N-gramok csoportosításával.

A tényleges jóslást egy ún. SVM (*Support Vector Machine* – SVM) végezte. Az SVM lényege, hogy bizonyos nyelvi jellegzetességekből (ez esetben a korábban képzett N-gramokból és topikokból) egy sokdimenziós (adott esetben több ezer dimenziós) térben egy vektort képeznek, ezzel reprezentálva a szöveg jelentését. Ahhoz, hogy a gép a „hasonló jelentés” (a vektorok egymáshoz képesti távolsága) alapján jóslni tudjon, egy halmaznyi ítélettel meg kell tanítani, hogy milyen jelentés esetén milyen végeredmény jön ki. Ehhez a gépnek egy halmaz ítéletet odaadnak, amelyben a végeredményt is megmondják (ez az ún. tanítóadat). Majd az adatok másik halmazáról megkérdezték a gépet, hogy mondjon predikciót, azaz sorolja be valamelyik csoportba a végeredményétől megfosztott ítéletet a szóhasználat alapján. A kísérlet végeredménye az volt, hogy a gép nagyjából 80 százalékos pontossággal volt képes, főként a tényleírás és a jogi érvelés szövegrészekre támaszkodva megjósolni az eredményt.

A kísérletet – amelynek további eredményei itt nem relevánsak – csak azért ismertettem, hogy képet kaphassunk arról, hogyan működik a predikció egy adott alkalmazásban. Mint láthatjuk, a predikció kétféle lehet: százalékos vagy bináris. A százalékos jóslás a fentebbi módszertől annyiban tér el, hogy egy adott szöveget egy 0-tól 1-ig terjedő skálán helyez el, így fejezve ki a per vagy a beadvány sikerességének esélyét. A rendszerek abban is különböznek, hogy mi alapján végzik el

¹⁵ Nikolaos Aletras et al.: Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective. *PeerJ Computer Science*, 2. (2016), e93.

a jóslást. Egyes rendszerekben csak a paramétereket lehet beállítani (például „alperes vagyok egy védjegyjogi perben, Y bíróság, Q bíró előtt, és X típusú beadványt akarok beadni, milyen az esélye a befogadásnak” – ilyen a *Premonition* rendszer *Motion analyzere*), a rendszer pedig valójában nem is predikciót mond, hanem a paraméterhalmaz alapján mutat egy százalékos eloszlást („Ennél a típusú beadványnál ebben a pertípusban, ezen bíró előtt 65 százalékos a sikeres befogadás.”). A másik típus, amikor valamilyen szövegmintázatot, („egy tényállást”) mutatnak a gépnek. Ez a gyakorlati alkalmazásokban általában az *e-Discovery* szoftverek része, amelyeknek nem is az elemzés a fő funkciójuk, hanem egy adott ügghöz tartozó összes dokumentum feltárása az ügyfél szerverein, levelezésében stb. Ilyenkor a szoftver ugyanazt a műveletsort végzi el, amelyet Aletras és társai cikkében láthattunk, azaz egy meghatározott szöveghalmazról mondja meg, hogy milyen nyerési esélyt mutat a bíróságon. Ez tehát azt az egészen szürreális funkciót teszi lehetővé ezekben a rendszerekben, hogy képesek egy elektronikus okirathalmaz alapján becslést készíteni a pereskedés esélyeiről.

A bíróelemző és ennek társadalmi hatásai

A bíróelemző programok tehát tulajdonképpen semmi mást nem tesznek, mint a fentebb ismertetett adatlegyűjtést egy bíróra vonatkozóan, azaz az eljáró bíró mint lekérdezési szempont alapján mutatják meg. Könnyen belátható, hogy minél nagyobb a dokumentumok granularitása, azaz minél több adatot tudnak kiemelni belőlük, annál nagyobb lesz az az „adatlap”, amelyet egy bíróról össze tudnak állítani.

Még azokból a jogrendszerekből is, ahol az ügyekből csak az ítélet áll rendelkezésre, rengeteg adat szedhető ki. Ezeket az adatokat három csoportba lehet sorolni:

- az ügy eljárási-technikai adatai (ügyszám, dátumok, előzményi és fellebbviteli kapcsolódó ítéletek, az eljáró bíróság, a társbírák, tanácsstagok személye, egyéb dátumok és nevek – a felek és képviselőik neve, esetleg más érintettek a perben);
- a jogi-tartalmi adatok (*content-oriented [meta]data*, amilyen az ügy vagy a nagyobb jogterület tárgya, a döntés iránya a perbeli szerepek szerint, a döntésben található anyagi jogi szerepkörök – például a bérbeadó-bérlő – szerint, a jogszabály- és jogeset-hivatkozások, beleértve a kvázi jogszabályokat is, mint az alkotmánybírói vagy Európai Emberi Jogi Bíróság ítéletei, a fellebbviteli ítélet iránya – helybenhagyás, megsemmisítés vagy megváltoztatás stb.);
- más, bonyolultabb NLP módszerekkel képzett adatok (BOW vagy más módszerekkel – lásd az előző oldalon – kiemelt adatok, szó-együttelfordulások stb. amelyekből a szöveg stílusára, egyedi jegyeire vagy más körülményekre lehet következtetni).

Ezekből az adatokból azután egy profil állítható össze, amely még következtetések nélkül is elég sok információt szolgáltat a bíról: például a bíró eddigi „portfólióját” (milyen ügyeket tárgyalt) milyen bíróságokon fordult meg, a személyi „együttállásokat” (kikkel tárgyalt eddig, és milyen ügyvédeknek milyen ítéleteket hozott, milyen bírákkal társasbíráskodott); a fellebbviteli sikerességét (mi az ítéleteinek a sorsa a felsőbíróságokon); az általa leggyakrabban idézett jogforrásokat, beleértve a korábbi precedenseket; milyen szerepköröknek milyen hányadban ítél (például inkább a bérbeadóknak vagy inkább a bérlőknek kedvez-e). Mindezt természetesen össze is lehet hasonlítani az egyes bírák szintjén, vagy például a teljes bíróság átlagához lehet hasonlítani.

A bíróelemző funkció még azokban az angolszász országokban sem örvend töretlen népszerűségnek, amelyekben a bírósági tárgyalások és ítéletek nyilvánossága jóval elfogadottabb, mint a kontinensen, a bírák ítéleteinek, későbbi idézésüknek¹⁶ vagy épp ideológiai hovatartozásuknak¹⁷ az elemzése pedig a jogi kultúra része. Nem csoda tehát, hogy amikor az első ilyen szolgáltatás Franciaországban megjelent, szinte azonnal hatalmas port kavart.

2019 júliusában több jogi technológiai portál¹⁸ is drámai körítéssel adta hírül, hogy Franciaországban „betiltották a bírák profilozását”, „kirúgták az adattudósokat a bíróságokról”, és ezzel jelentősen csökkentették a bíróságok transzparenciáját.¹⁹ Az adatalapú jogi elemzéssel foglalkozó legismertebb francia cég tulajdonosa²⁰ „a demokrácia szégyenének”²¹ nevezte a döntést.

A jogszabály,²² amelyet még az év tavaszán fogadott el a francia parlament, a bírósági reform részeként született, és természetesen nemcsak a tiltásról szól, hanem a korábbinál jóval szélesebb – lényegében teljes – körben elrendeli a francia bíróságok ítéleteinek anonimizált közzétételét a világhálón. Csakhogy a közzététellel egy időben beiktat egy olyan szabályt is, amely szerint a bírák és a bírósági dolgozók munkájának „értékelésére, elemzésére és összehasonlítására” nem lehet felhasználni a közzétett dokumentumokban található adatokat. A szabály megszegőit büntetni

¹⁶ Montgomery N. Kosma: Measuring the Influence of Supreme Court Justices. *The Journal of Legal Studies*, 27. (1998), 2. Part 1. 333–372.

¹⁷ Például Youngsik Lim: An Empirical Analysis of Supreme Court Justices’ Decision Making. *The Journal of Legal Studies*, 29. (2000), 2. Part 1. 721–752.

¹⁸ Simon Taylor: French Data Analytics Law Won’t Stop Analytics, Attorneys Say. [Law.com](https://www.law.com), 2019. június 7.

¹⁹ John O. McGinnis: Transparency and the Law in France. *Law & Liberty*, 2019. június 20.

²⁰ A Predictice weboldala.

²¹ Jason Tashea: France Bans Publishing of Judicial Analytics and Prompts Criminal Penalty. *ABA Journal*, 2019. június 7.

²² LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice Article 33.

rendeli a francia Btk.-ban is megtalálható „személyes adattal való visszaélés”²³ bűncselekménye miatt.

A *Vervassungsblogon* megjelent egyik írás²⁴ a szigor hátterében egy korábbi ügyet vél felfedezni: egy Michaël Benesty nevű jogász-informatikus még 2016-ban publikált egy adatelemzésen és gépi tanuláson alapuló írást²⁵ a francia bíróságok menekültügyi gyakorlatáról. Ebből kiderült, hogy az egyes bíróságok és bírák ítéletei közt semmivel nem indokolható különbségek vannak. A francia bíróságok gyorsan reagáltak: részint tagadták a vádakat, részint pedig megkérdőjelezték Benesty módszertanát.

Az új jogszabálynak természetesen akadtak védelmezői is. A Lorraine-i Egyetem jogászprofesszora például arról ír²⁶ az egyik nagy kereskedelmi kiadó oldalán, hogy az angolszász sajtó reagálása egyszerű félreértés, amely a *common law* és a francia jog közti különbségből ered. Franciaországban a bírák nem a saját nevükben hozzák az ítéletet, hanem a köztársaság nevében, és máig tartják magukat a logikai szillogizmus (szükségszerű következtetés) fikciójához. Ráadásul a bírák sok esetben tanácsban ítélik meg, amelyeken belül az egyes bírák álláspontját nem láthatjuk. A bírák profilját tehát így nehéz felállítani, illetve azon mindezek a körülmények sokat torzíthatnak. Végül, mivel a jogszabály kizárólag a bírák és a bírósági dolgozók név szerinti elemzését, összehasonlítását és rangsorolását tiltotta meg, másféle adatalapú elemzések továbbra is lehetségesek. Az ítéleteket, ítéletek különböző szempont szerinti csoportjait, bíróságokat, sőt akár tanácsokat továbbra is lehet elemezni, összehasonlítani. Az ügyek lehetséges jövőbeli kimeneteiről továbbra is lehet előrejelzéseket mondani.

Lehetőségek és aggályok a big data alapú elemzőprogramokkal kapcsolatban – az elemzőprogramok jövője

A francia eset jól mutatja az adatalapú jogi elemzéssel kapcsolatos dilemmákat, amelyek a big data személet jogban történő megjelenése óta szinte folyamatosan napirenden vannak. Ezek között vannak olyan aggályok, amelyek általánosak, nem függenek jogrendszerétől, és vannak olyanok, amelyek kifejezetten azt róják fel, hogy e szoftverek az angolszász (*common law*) jogi kultúrához kötődnek, és kontinentális jogi környezetben nem használhatók. A továbbiakban ezt a két érvcsoportot

²³ Code pénal Article 226–18.

²⁴ Malcolm Langford – Mikael Rask Madsen: France Criminalises Research on Judges. *Vervassungsblog*, 2019. június 22.

²⁵ Michaël Benesty: L’Impartialité de certains juges mise à mal par l’intelligence artificielle. *Village de la Justice*, 2016. március 25.

²⁶ Florence G’Sell: Predicting Courts’ Decisions is Lawful in France and Will Remain so. *Actualités du droit*, 2019. július 2.

vizsgálom meg. Mielőtt azonban rátérnék az aggályokra röviden rekapitulálok az adatalapú jog híveinek érveit, főként Daniel Katz sokat idézett cikke²⁷ alapján.

Az adatalapú jog előnyei

Katz kiinduló érve az, hogy a jogi szakmák – különösen az ügyvédi szakma – rengeteg predikciós elemet tartalmaz. Van-e ebben az ügyben esélyünk? Hogyan fog ítélni a bíróság? Mennyibe fog ez nekem kerülni? Milyen kockázatai vannak ennek vagy amannak a szerződéses rendelkezésnek, vagy épp annak, ha kihagyjuk a szerződésből?²⁸ Mindezeket a kérdéseket egy ügyvédnek naponta meg kell tudni válaszolni, de támaszkodni csak a saját múltbeli tapasztalataira és intuícióira tud. Az adatalapú elemzésnek azonban nincsenek ilyen korlátai. Egy-egy jogterületnek szinte minden releváns ítéletére rálát, és olyan nagy mennyiségű adatot tud elemezni, amennyit egy ember képtelen. Nem elfogult – és nem fárad.

Katz először a pereskedés költségeit vizsgálja meg. A TyMetrix²⁹ rendszer példáját hozza fel, ahol egy-egy jogterületre vonatkozó perek adatai alapján lehet lekérdezni, hogy milyen jogi költségekre lehet számítani. Második példája az ügyvédek hatékonyságának összehasonlítási lehetősége.³⁰ Ez a peres adatokból egyértelműen kiolvasható: látható, hogy ki volt az eljáró ügyvéd, és hogy megnyerte-e a pert. És végül persze a legfontosabb: az esetek kimenetelét is meglehetősen pontossággal lehet jósolni – mondja Katz. Három területen elemzi a prediktív rendszereket, az amerikai Legfelsőbb Bíróság (LB), a Szabadalmi Bíróság és a tőkepiaci csalásokkal kapcsolatos csoportos igényérvényesítések (*securities fraud class action*) területén.

Érdekes közelebbről is megnézni az LB-vel kapcsolatos példáját, amely egy 2002-ben lefolytatott kísérlet.³¹ Ebben a kísérletet végzők az USA Legfelsőbb Bírósága által az adott évben kibocsátott összes ítélet kimenetelét prediktálták gépi úton, az ügyekből képzett különböző változókból. A kísérlethez szakértő jogászokból álló kontrollcsoportot is használtak. A gép 75, míg a szakértői kontrollcsoport 59 százalékos pontosságot ért el. A gépi predikció a múltbeli adatokból kiszedett változókat (mint például az ügy témája és a bíró politikai irányultsága³²) vetítette

²⁷ Katz (2013): i. m.

²⁸ Katz (2013): i. m. 928.

²⁹ A TyMetrix egy SaaS rendszerű (Software as a Service, online szoftver) e-számlázó és ügymenedzsment megoldás, amelynek valóban van predikciós modulja. A rendszer jelenleg a Wolters Kluwer tulajdonában van (TyMetrix 360° Application Modules).

³⁰ Ilyen modulja van a Premonition rendszernek.

³¹ Theodore W. Ruger et al.: Essay, The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-making. *Columbia Law Review*, 104. (2004), 4. 1150–1209.

³² Ezek a változók a következők: (1) a panasz körzete; (2) az ügy tárgyának besorolása; (3) a kérelmező típusa (például szövetségi kormányzat, munkáltató); (4) az alperes típusa; (5) az alsóbíróság

a jelen ítéleteire, és ezek alapján mondott jóslatot. Meghökkenítő, de az ügy tényállása, jogi érvei ebben a modellben nem is játszottak szerepet. A gép hat egyszerű, és nem is „jogi-szakmai” változó alapján sokkal jobban jóslott, mint a szakértők.

Általános aggályok

Devins és szerzőtársai szerint a big data alapú elemzésekkel kapcsolatban három ellenérvet lehet felhozni a jog területén.³³

Az első érv: semmilyen, még a legobjektívebbnek tűnő adathalmaz és -elemzés, következtetés sem lehet teljesen „objektív” és „értékmentes”, amit az ilyen szolgáltatások gyakran állítanak magukról. Egyrészt minden megfigyelés „elmélettel terhelt” (*theory-laden*), különösen az olyan komplex rendszereknél, mint a jog. Már az is egy előzetes elméletet igényel ugyanis, hogy egyáltalán miről gyűjtünk adatot. Másrészt az adatok értelmezése további elméletet, keretezést, értékválasztást, tehát egy sor olyan mozzanatot igényel, amelytől tovább veszít objektivitásából. Ehhez az érvhöz még hozzátehetjük, hogy az elméletek pedig értékekkel terheltek, az értékekre, illetve ezek jogi lecsapódásaira, az elvekre pedig különösen jellemző, hogy egymással is ellentmondásba kerülhetnek. Ez azt jelenti a big data alkalmazásokra nézve, hogy ugyanazon nyelvi mintázat vagy adathalmaz mind az elméleti keretezése, mind valamilyen értékkontextusba történő beleágyazódása miatt teljesen eltérő eredményekre vezethet, amelyet a gép nem biztos, hogy érzékel.

A második érv az adatalapú jog ellen az, hogy annak legfőbb erénye, a jósló ereje egyszerűen illuzórikus, és voltaképpen semmire sem használható. A jog nem determinisztikus rendszer, és ráadásul állandóan ki van téve a változásnak. Mindig vannak olyan szereplői, akik váratlan lépéseket tesznek, váratlan érvelést alkalmaznak, amely adott esetben sikerre vezet. A százalékos valószínűség pedig semmit nem mond a konkrét ügy kilátásairól, hiszen nem oksági alapon nyugszik. A jogban az oksági összefüggéseknek pedig nagy szerepük van.

Ez az érv valójában két érvet tartalmaz. Az érv első fele arról szól, hogy a jogi eljárások kognitív harcok, nyelvi „csaták”, amelyet nemcsak a csatában elhangzó konkrét adatok, jogszabály-hivatkozások stb. döntenek el, hanem mindezek meggyőző ereje, azaz a jogász érvelés minősége is. A jogász érvelés pedig két szinten is jelentkezik: egyfelől a jogszabályok értelmezésének, másfelől a tények konstruálásának területén.³⁴ Mindkét terület igen nagy teret hagy az érvelési „ügyességnek”.

ítéletének ideológiai iránya (liberális vagy konzervatív); (6) érvelt-e a kérelmező azzal, hogy a jogszabály vagy a gyakorlat alkotmányellenes.

³³ Caryn Devins et al.: The Law and Big Data. *Cornell Journal of Law and Public Policy*, 27. (2017), 2. 357–414.

³⁴ Zsolt Zódi: Algorithmic Explainability and Legal Reasoning. *The Theory and Practice of Legislation*, 10. (2022), 1. 67–92.

Ez az oka annak is, hogy az „algoritmusok megmagyarázhatósága” és egy „jogi döntés” megmagyarázása, azaz a jogi érvelés, a *legal reasoning* olyan drámai módon különbözik egymástól. A jogi érvelésben a tények megfelelő módon történő elmesélése, tálalása, a hétköznapi narratíváink, az életszerűség nagyon fontos szerepet játszik, ahogy a jogszabályok nyitott szövedékű szövegeinek értelmezése is,³⁵ sőt, voltaképp a jogi döntés nem más, mint e két művelet egymásra tekintettel történő végzése. A gépi elemzések azonban kemény tényeken és kemény szabályokon alapulnak. A tények nem narratívák, hanem kizárólag adatok, a szabályok pedig nem többféleképp értelmezhető hétköznapi nyelvi objektívációk, hanem algoritmusok.

Az érv második fele szintén egy filozófiai (jóllehet nem episztemológiai, hanem ontológiai) érvet tartalmaz, mégpedig azt, hogy a statisztikai jellegű összefüggések (például a korreláció), amelyek a valószínűségi összefüggésekkel dolgoznak, a jogban azért nem működnek jól, mert a jog mindig egyedi esetekkel foglalkozik, és nem érdeklik a statisztikai összefüggések. Hogy történt mindez ebben a konkrét esetben? – kérdezi a jogász. És az, hogy egy bizonyos ügycsoportban 90 százalékban pert vesztenek a felek, nem mond semmit arról, hogy benne lesznek-e a 10 százalékban.

Végül a harmadik érv, amelyet Devins és szerzőtársai felhoznak az adatalapú jog ellen: annak bizonyos területeken való alkalmazása kifejezetten veszélyes lehet. Itt a jól ismert aggályokat olvashatjuk az algoritmusokba és az adatokba „belekódolt” diszkriminációval kapcsolatban, a jog megmerevedésének rémét, mivel az adatalapú jog mindig csak a múltat nézi.

A common law – kontinentális jog különbségéből fakadó aggályok

Az elemző programok angolszász területen jöttek létre. A kontinensen nemcsak azért nehéz ilyen szolgáltatásokat létrehozni, mert eltérő a mögöttes jogi kultúra, hanem azért is, mert nem állnak rendelkezésre olyan mennyiségben adatok (dokumentumok, iratok), mint például az USA-ban. Ahogy azt a Lex Machina rendszer ismertetésénél már jeleztem, az feldolgozza a PACER rendszerben található ügyiratokat. Ez – leszámítva azokat az eseteket, amikor zárt iratkezelést kértek, vagy az ügy természete miatt nem nyilvánosak az iratok, a teljes iratállományt jelenti: a felek beadványaival, tárgyalások jegyzőkönyveivel, szakértői véleményekkel. Könnyen belátható, hogy mennyivel több információ nyerhető ki ezekből az iratokból, mint a kontinentális jogrendszerekben egyedül nyilvános – anonimizált – ítéletekből. Ehhez még azt is hozzá kell tenni, hogy egy sor országban

³⁵ H[erbert] L. A. Hart: *A jog fogalma*. Budapest, Osiris Kiadó, 1995. 147–159.

ráadásul nem is a teljes ítélethalmazt – sőt még csak nem is egy reprezentatív részét – publikálják, hanem annak egy kis töredékét.³⁶

Teljesen más az adatvédelem rendszere is a két kultúrában. Európában sokkal magasabbak az adatvédelmi sztenderdek. Ha csak azt az egyszerű tényt nézzük, hogy a *common law* rendszerekben a felek nevéől nevezik el az ügyet, már ebből azonnal látható, mekkora a különbség a két jogi kultúra között. A létező és jogos adatvédelmi korlátokat ráadásul olykor szinte érthetetlen pluszakadályokkal is nehezítik, mint például hogy a gazdasági perekben a cégneveket is ki kell törölni az ítéletekből, jöllehet a cégeket nyilvánvalóan nem védi az adatvédelem.

Talán még ennél is fontosabb, hogy az USA-beli rendszerek egyik fő felhasználási területe a *forum shopping*, azaz a felek a statisztikák és az adatok fényében döntenek el, hogy hol kezdeményeznek eljárást. Erre a legtöbb kontinentális jogrendszerben a szigorú hatásköri és illetékességi szabályok miatt nincs lehetőség. Ugyanakkor verseny zajlik a világ jogrendszerei közt, és az adatalapú elemzések a fórumok megválasztásának kérdését is új megvilágításba helyezik.

A kulturális különbség azonban még lényegesebb. A francia érvelés fontos része, hogy a bíró tehát „gép”, amely mindig ugyanúgy működik, nehéz tehát a „profilját” és a „személyiségét látni”. A bíróelemzők azonban élesen mutatnak rá, hogy a jog alapvetően retorikai teljesítményeken alapszik, a jogalkalmazást szillogizmusként bemutató modell még eszményként sem működik. Egyelőre nehezen jósolható meg, hogy milyen következményei lesznek annak, hogy a kontinentális jogrendszerekben egyszer s mindenkorra le kell számolni ezzel az eszményképpel.

Merre tovább, elemzők?

A francia betiltás fogadtatása jól mutatta, hogy még a részleges korlátozás sem járható út: ha van adat, akkor ezek a szolgáltatások, beleértve a bíróprofilozást is, továbbra is velünk maradnak, legfeljebb „otthon, a négy fal közt” fogják csinálni azok, akik ettől remélik, hogy előnyre tesznek szert egy eljárásban. Ezt persze lehet skeptikusan vagy szomorú beletörődéssel kezelni, de sokkal jobb, ha inkább az előnyeiket nézzük, illetve azt, hogy mire használhatók ezek a szolgáltatások, és mire nem érdemes őket használni.

A legfontosabb, hogy úgy tűnik, a szolgáltatásokat ellenzők valójában összekevernek két dolgot, a big data alapú *jogi döntés* és a big data alapú *tanácsadás*, valamint *elemzés* tevékenységeit. Attól, hogy egy bíró vagy egy rendszer nem dönthet statisztikai alapon, egy ügyvéd miért ne nézhetné meg a statisztikáját egy bírónak vagy egy jogterületnek, miért ne láthatná a főbb érveket, az ítélkezé-

³⁶ A magyar peres outputnak nagyjából 2 százalékát publikálják. Zódi Zsolt: *A korábbi esetekre történő hivatkozások mintázatai a magyar bíróságok ítéleteiben*. MTA Law Working Papers 2014/01. Budapest, Magyar Tudományos Akadémia Társadalomtudományi Kutatóközpont, 2014. 8. (2. táblázat).

irányait? Ezt az első általánosabb következtetést tehát úgy lehet megfogalmazni, hogy az adatalapú elemzéseknek az ügyvédi munka egyik elemeként igenis lehet létjogosultságuk.

A bírák számára is lehetnek az ilyen szolgáltatásnak előnyei, hiszen összeszedve láthatják magukról azokat az információkat, amelyek egyébként nyilvános adatbázisokból származnak, és ez segíthet adott esetben abban, hogy önmagukat, saját rossz beidegződéseiket, eddig fel nem ismert szokásaikat, sztereotip szóhasználatukat stb. jobban lássák, és javíthassanak rajtuk.

A gépi elemzések segítenek felfedni a bírói gyakorlat egyenetlenségeit is. Sokan erre azt hozzák ellenérvként, hogy nincs két egyforma ügy, és „biztosan oka van annak”, ha egy bíró egy látszólag hasonló ügyben ellentétesen dönt. Ugyanakkor látni kell, hogy az ilyen – esetleg látszólagos, de potenciálisan valóságos – egyenetlenségek is könnyebben előkerülnek gépi elemzéssel. Az pedig, hogy a felszínre hozott adatokkal mit kezdünk, már rajtunk múlik.

Előre látni a jövőt, és ez alapján jobb döntéseket hozni persze a végső cél, de eggyel közelebbi cél lehet csak az is, hogy bizonyos jelenségeket egyszerűen jobban megértsünk a „kemény tények” alapján. Ha egy bíróról minden, az ítéleteiből kiolvasható adatot látunk, akkor sokkal árnyaltabb képet kapunk róla még akkor is, ha nem akarunk az adott ügyben jóslást készíteni. Egy jogterület bírói gyakorlatának áttekintését és megértését nagyban elősegíti, ha látjuk, milyen hosszúak a perek, kik a területen a legtöbbet tárgyaló bírák, milyen arányban hagyják másodfokon helyben az elsőfokú ítéleteket, és így tovább. Különösen nagy segítség lehet a jogtudomány (az ún. dogmatikai tudomány) számára az adatalapú elemzés mint kontrolleszköz.³⁷

A legtöbb szolgáltató (így például a Premonition is³⁸) azzal érvel, hogy aki ilyen elemzőket használ, az a jogi kognitív harcban előnyre fog szert tenni. Csakhogy ez az előny mindössze addig tart, ameddig a másik fél elő nem fizet ugyanerre a szolgáltatásra. Ezért érdemes néhány szót ejteni arról is, hogyan fog kinézni a jogrendszer, a jogi munka, amikor mindenki számára rendelkezésre állnak majd ezek az adatok. Ezt a kérdést érdemes kétfelé bontani, egyrészt azokra az adatokra, amelyek egy percsoporthoz vagy bíróság aggregált adatait vagy az ezekből levont következtetéseket mutatják, és azokra, amelyek emberek (bírák vagy jogi képviselők) teljesítményét vagy szakmai előéletét mutatják.

Az első típus, a nem személyhez kötődő adatok véleményem szerint elsősorban úgy fognak hatni, hogy a pereskedés kilátásait megjósolva elsősorban a peren kívüli megegyezéseket fogják elősegíteni (ahogy erre már vannak is jelek). A nagyvállalatok a jogi költségeik tervezhetősége miatt ma is használják a jogtanácsosaikat vagy az ügyvédek – legtöbbször százalékos formában kifejezett – előrejelzéseit például a céltartalékképzéshez (a per elvesztésének lehetséges költségeit

³⁷ Lásd Zódi (2020): i. m. 109.

³⁸ A Premonition weboldala.

megképezve céltartaléknak) vagy a peren kívüli megegyezés során felajánlott összeg kalkulálásához. Az aggregált formában, szisztematikusan prezentált érvek a viszonylag egyszerű ügyekben hasznos segítséget adhatnak az ügyvédnek a pertaktika kidolgozásához, vagy amennyiben a statisztikailag „vesztes” oldalt képviselik, új érvek kidolgozásához. Emiatt az elemzők által „aggregált múlt” nem feltétlenül fojtja meg a jogi innovációt, mert a joggyakorlat szisztematikus, adatalapú prezentációjával éppenséggel a jogi innovációt is segíthetik.

Más a helyzet a jogi képviselők és a bírák elemzésével. A „win-rate” kiszámolásának és mutatójának egy ügyvéd esetében igen komoly presztízsszomboroló hatása lehet, így ezzel nagy valószínűséggel ugyanaz fog történni, mint sok más hasonló esetben (például a hitelminősítés esetén): ki fognak alakulni az egységes módszerek, amelyek alapján ilyen információt közölni lehet valakiről, és ezek a módszerek egyre transzparenssebbek lesznek. Még azt sem tartom kizártnak, hogy jogszabályi szinten fogják előírni, hogyan kell kiszámítani egy ügyvéd win-rate-jét, milyen egyéb információkat kell közölni ennek kapcsán, és milyen helyzetekben nem lehet ezeket a mutatókat használni (például ha nincsen elegendő mennyiségű adat). A bírák értelemszerűen a saját magukra vonatkozó elemzést ismerni fogják, és amennyiben az nincsen az ínyükre, tenni fognak ellene, főként úgy, hogy tudatosan a róluk kialakult adatalapú kép ellenében cselekednek, viselkednek (ha például az az értékelés, hogy „ez a bíró mindig a bankok javára ítél” megpróbál jobban odafigyelni, hogy kiegyensúlyozottabb legyen az ítélezése stb.). Az azonban biztos, hogy ezekkel a szolgáltatásokkal kapcsolatban az újdonság varázsa hamar el fog múlni, és majd belesimulnak a hagyományos jogászkodás mindennapjaiba, ahogy az elektronikus jogi forráskutatás vagy a Google-kereső használata esetében is történt.

Felhasznált irodalom

- Aletras, Nikolaos – Dimitrios Tsarapatsanis – Daniel Preoțiuc-Pietro – Vasileios Lamos: Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective. *PeerJ Computer Science*, 2. (2016), e93. Online: <https://doi.org/10.7717/peerj-cs.93>
- Ashley, Kevin D.: Reasoning with Cases and Hypotheticals in HYPO. *International Journal of Man-Machine Studies*, 34. (1991), 6. 753–796. Online: [https://doi.org/10.1016/0020-7373\(91\)90011-U](https://doi.org/10.1016/0020-7373(91)90011-U)
- Benesty, Michaël: L’Impartialité de certains juges mise à mal par l’intelligence artificielle. *Village de la Justice*, 2016. március 25. Online: www.village-justice.com/articles/impartialite-certains-juges-mise,21760.html
- Best Legal Analytics Software 2021. Online: www.360quadrants.com/software/legal-analytics-market
- Code pénal Article 226–18. Online: www.legifrance.gouv.fr/affichCodeArticle.do?idArticle=LEGIARTI000006417968&cidTexte=LEGITEXT000006070719

- Devins, Caryn – Teppo Felin – Stuart Kauffman – Roger Koppl: The Law and Big Data. *Cornell Journal of Law and Public Policy*, 27. (2017), 2. 357–414.
- G'Sell, Florence: Predicting Courts' Decisions is Lawful in France and Will Remain so. *Actualités du droit*, 2019. július 2. Online: www.actualitesdudroit.fr/browse/tech-droit/donnees/22630/predicting-courts-decisions-is-lawful-in-france-and-will-remain-so
- General Guidelines 2020. (for Search Quality Rating of Google Search Engine). 2021. október 19. Online: <https://guidelines.raterhub.com/searchqualityevaluatorguidelines.pdf>
- Hart, H[erbert] L. A.: *A jog fogalma*. Budapest, Osiris Kiadó, 1995.
- Hoge, Patrick: Lex Machina: 'Moneyball' Meets Patent Lawsuits. *San Francisco Business Times*, 2013. szeptember 3. Online: www.bizjournals.com/sanfrancisco/blog/2013/09/lex-machina-mines-lawsuit-data.html?page=all
- Katz, Daniel Martin: Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry. *Emory Law Journal*, 62. (2013), 4. 909–966.
- Kosma, Montgomery N.: Measuring the Influence of Supreme Court Justices. *The Journal of Legal Studies*, 27. (1998), 2. Part 1. 333–372. Online: <https://doi.org/10.1086/468023>
- Langford, Malcolm – Mikael Rask Madsen: France Criminalises Research on Judges. *Verfassungsblog*, 2019. június 12. Online: <https://verfassungsblog.de/france-criminalises-research-on-judges/>
- LexisNexis Acquires Premier Legal Analytics Provider Lex Machina. *Lex Machina*, 2015. november 23. Online: <https://lexmachina.com/media/press/lexisnexis-acquires-lex-machina/>
- Lim, Youngsik: An Empirical Analysis of Supreme Court Justices' Decision Making. *The Journal of Legal Studies*, 29. (2000), 2. Part 1. 721–752. Online: <https://doi.org/10.1086/468091>
- LOI n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice Article 33. Online: www.legifrance.gouv.fr/eli/loi/2019/3/23/2019-222/jo/article_33
- Mayer-Schönberger, Viktor – Kenneth Cukier: *Big Data. The Essential Guide to Work, Life, Learning in the Age of Insight*. London, John Murray, 2013.
- McGinnis, John O.: Transparency and the Law in France. *Law & Liberty*, 2019. június 20. Online: www.lawliberty.org/2019/06/20/transparency-and-the-law-in-france/
- Mitchell, Ryan: *Web Scraping with Python*. Sebastopol (CA), O'Reilly Media, 2018.
- Rapoport, Nancy B. – Joseph R. Tianio Jr.: Legal Analytics, Social Science, and Legal Fees: Reimagining Legal Spend Decisions in an Evolving Industry. *Georgia State University Law Review*, 35. (2019), 4. 1269–1304.
- Ruger, Theodore W. – Pauline T. Kim – Andrew D. Martin – Kevin M. Quinn: Essay, The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-making. *Columbia Law Review*, 104. (2004), 4. 1150–1209. Online: <https://doi.org/10.2307/4099370>
- Tashea, Jason: France Bans Publishing of Judicial Analytics and Prompts Criminal Penalty. *ABA Journal*, 2019. június 7. Online: www.abajournal.com/news/article/france-bans-and-creates-criminal-penalty-for-judicial-analytics
- Taylor, Simon: French Data Analytics Law Won't Stop Analytics, Attorneys Say. *Law.com*, 2019. június 7. Online: www.law.com/legaltechnews/2019/06/07/french-data-analytics-law-wont-stop-analytics-397-21251/?slreturn=20191006132927
- Zódi Zsolt: A jogi szövegek mint big data. In Jakab András – Sebők Miklós (szerk.): *Empirikus jogi tanulmányok*. Budapest, Osiris Kiadó – Magyar Tudományos Akadémia Társadalomtudományi Kutatóközpont, 2020. 93–109.

- Zódi Zsolt: *A korábbi esetekre történő hivatkozások mintázatai a magyar bíróságok ítéleteiben*. MTA Law Working Papers 2014/01. Budapest, Magyar Tudományos Akadémia Társadalomtudományi Kutatóközpont, 2014. 1–55.
- 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>
- Zódi Zsolt: Jog és jogtudomány a big data korában. *Állam- és Jogtudomány*, 58. (2017), 1. 95–114.