



**MESTERSÉGES INTELLIGENCIA**  
Nemzeti Laboratórium

Társadalomtudományi  
Kutatóközpont 



# A Mesterséges Intelligencia Nemzeti Laboratóriumhoz kapcsolódó kutatások a Társadalomtudományi Kutatóközpontban (2020-2023)

Társadalomtudományi Kutatóközpont  
2023



*Tisztelettel köszöntöm a Társadalomtudományi Kutatóközpont új kiadványának olvasóit! Ezen kötet a Mesterséges Intelligencia Nemzeti Laboratóriumon belüli kutatási témáit és eredményeit mutatja be.*

*Az MI alkalmazásának széleskörű elterjedése és a technológiák használatának a mindennapi életre gyakorolt hatásai miatt az MI társadalmi hatásainak kutatása is egyre fontosabb. Ezért a Társadalomtudományi Kutatóközpont részvétele kiemelt fontosságú a Mesterséges Intelligencia Nemzeti Laboratórium munkájában. A TK kutatói alkalmazzák is a mesterséges intelligenciát mint technológiát az aktuális társadalmi, jogi, foglalkoztatási és politikai vonatkozású kutatásaikban. Ehhez a Mesterséges Intelligencia Nemzeti Laboratórium tud szakmai keretet biztosítani. Biztos vagyok benne, hogy mindenki talál izgalmas és érdekfeszítő témákat ebben a kiadványban.*



**Benczúr András,**  
a Mesterséges Intelligencia Nemzeti Laboratórium  
szakmai vezetője



*Örömmre szolgál e kötet megjelenése, amely bepillantást enged abba a sokféle munkába, amelyet a Társadalomtudományi Kutatóközpont munkatársai végeznek a Mesterséges Intelligencia Nemzeti Laboratóriumában. A külső szemlélőnek első pillantásra talán nem evidens, hogy mi köze a társadalomkutatónak a mesterséges intelligenciához. De nem szabad elfelejteni, hogy minden technológia kérdés végső soron társadalmi ügy: az adott társadalmi, gazdasági, kulturális kontextusban működik (vagy nem); használata társadalmi, vagyis attitudinális, viselkedésbéli, kulturális, vagy intézményi változásokat indít el vagy feltételez; etikai és jogi dilemmákat és új szabályozási igényt vet fel; gazdasági hasznokat generál egyes szereplőknél, míg kihívást intéz más vállalatok számára; munkahelyeket hoz létre és szüntet meg; és így tovább. Mindehhez tegyük hozzá, hogy a mesterséges intelligenciára alapozott módszerek bekerültek a társadalomkutatók eszköztárába is. Sőt, büszkén mondhatom, hogy kollégáim nem csupán használják ezeket a módszereket, hanem fejlesztésükhöz is hozzájárultak. Jelen kötet tehát a mesterséges intelligencia használatának társadalmi aspektusaiba enged betekintést, felvillantva e problémakör tematikus gazdagságát.*

**Boda Zsolt,**  
a Társadalomtudományi Kutatóközpont főigazgatója





Ságvári Bence,  
a TK MILAB szakmai társ-vezetője



Néhány évtizeddel ezelőtt kevesen gondolták, hogy a mesterséges intelligencia és a társadalomtudomány milyen sok területen és szerepben találhat egymásra. Napjainkban azonban már senki sem kapja fel a fejét azon, ha mérnökök, természettudósok és társadalomtudósok használnak a kutatásaikban mesterséges intelligencia-alapú kifinomult elemzési eszközöket, vagy közösen azon dolgoznak, hogy hatékony, a társadalom számára hasznos, etikus és igazságos MI-alapú technológiák jöhessenek létre. A Társadalomtudományi Kutatóközpontban az elmúlt években elindult folyamatok eredményeként mind az alkalmazott módszertanokban (természetes nyelvfeldolgozás, hálózatok, big data, adatdonáció stb.), mind pedig a kutatási témákban megjelent a mesterséges intelligencia. Kialakult egy olyan kutatói közösség, amelynek tagjai képesek a társadalomtudományok elméleti hagyományait ezen az új technológiai szemüvegen keresztül szemlélni. Őszintén reméljük, hogy az intézetünkben folyó MI-kutatások hozzájárulhatnak a magyarországi tudományok fejlődéséhez, a társadalmi jólét és jóllét növekedéséhez, és nem utolsósorban, a TK tovább növelheti jelenlétét a nemzetközi tudományos életben.

*A mesterséges intelligencia 21. századi társadalmi, gazdasági, politikai hatásai minden túlzás nélkül fordulópontot jelentenek az emberiség történetében. Olyan forradalom zajlik, mely jelentőségében a korábbi ipari forradalmakhoz hasonlítható: ahogy a gőzgép vagy az elektromos energia használatának elterjedése régi foglalkozások eltűnéséhez és új munkahelyek megjelenéséhez vezetett, úgy rendezi át a mesterséges intelligencia korunk munkaerőpiacát. E hatások között egyszerre vannak pozitív és negatív fejlemények: a testreszabott szolgáltatások és applikációk milliárdnyi felhasználót kiszolgáló megjelenése együtt jár a személyes adatok védelmének sérülésével, illetve extrém esetekben politikai vagy üzleti célú manipulációra való felhasználásával. A Társadalomtudományi Kutatóközpont a Mesterséges Intelligencia Nemzeti Laboratórium egyik alapító tagjaként azt a célt tűzte ki maga elé, hogy új, például nyelvtechnológiai alkalmazások fejlesztése mellett elemezze az új technológia komplex gazdasági-társadalmi-politikai hatásait. Ebben együttműködik sok tucatnyi hazai és nemzetközi intézménnyel, interdiszciplináris kutatási projekteket visz, több száz résztvevőt megmozgató oktatási programot, illetve ismeretterjesztési előadássorozatot indított. A jelen kiadvány célja, hogy körképet adjon a Kutatóközpont kiterjedt tudományos tevékenységéről a mesterséges intelligencia kutatása kapcsán.*

Sebők Miklós,  
a TK MILAB szakmai társ-vezetője



# A TK MILAB kutatási projektjeinek áttekintése

|       | A projekt elnevezése                                                                  | Projektvezető               | Munkacsomag                 |
|-------|---------------------------------------------------------------------------------------|-----------------------------|-----------------------------|
| 8–9   | MI attitűdök és narratívák                                                            | Ságvári Bence               | Szenzor, IoT, távközlés     |
| 10–11 | poltextLAB – Politikai és jogi szövegbányászat mesterséges intelligencia segítségével | Sebők Miklós                | Nyelvtechnológia            |
| 12–13 | Octopus Research Tools (ORT)                                                          | Ságvári Bence               | Szenzor, IoT, távközlés     |
| 14–16 | A mesterséges intelligencia jogi kihívásai                                            | Mezei Kitty                 | Biztonság, személyes adatok |
| 17    | A morális pánikgomb (MPG) modellezése hagyományos és NLP modellekkel                  | Sik Endre                   | Nyelvtechnológia            |
| 18    | Jegybanki kommunikáció mérése nagy nyelvi modellekkel                                 | Máté Ákos                   | Nyelvtechnológia            |
| 19    | Magyar nyelvű benchmark adatbázisok gépi tanulás fejlesztésére                        | Kiss László<br>Molnár Csaba | Nyelvtechnológia            |
| 20–21 | Meddig érvényes egy klasszifikáció? – Oltásellenes kommentek klasszifikációja         | Kmetty Zoltán               | Nyelvtechnológia            |
| 22–23 | Politikai szövegek véleményelemzése                                                   | Ring Orsolya                | Nyelvtechnológia            |
| 24    | Szöveges kutatási források tematikus feltárása a TK adatrepozitóriumaiban             | Gárdos Judit                | Nyelvtechnológia            |
| 25    | Társadalmi-kognitív hatások a jogi döntéshozatalban: formális modellezés              | Péli Gábor                  | Nyelvtechnológia            |
| 26    | Mesterséges intelligencia és drónhasználat a mezőgazdaságban                          | Kovács Imre                 | Társadalmi és jogi hatások  |
| 27    | Munka 5.0 – Technológia és munka a mesterséges intelligencia korában                  | Illésy Miklós               | Társadalmi és jogi hatások  |
| 28    | Oksági gépi tanulás: az alapelvektől a datasheet-ig                                   | Janky Béla                  | Társadalmi és jogi hatások  |





## MI attitűdök és narratívák

Projektvezető: Ságvári Bence

A mesterséges intelligencia (MI) fejlődése az életünk szinte minden területére hatással lesz. A munka világa, a közlekedés, az egészségügy, az oktatás, az igazságszolgáltatás és a nemzetbiztonság mind olyan területek, ahol lassan, de biztosan megjelennek azok a technológiák és eljárások, amelyek eredményeként egyre több döntést vagy döntés előkészítést fogják MI-n alapuló rendszerek elvégezni.

Alapvető kérdés ugyanakkor, hogy mindez mennyire elfogadott a társadalom részéről. Az elkövetkezendő időszak fejlődését döntően fogja meghatározni, hogy az emberek mennyire vannak tisztában a technológia kínálta lehetőségekkel, mit gondolnak ezeknek a megoldásoknak a „világjobbító” szándékairól, vagy éppen mennyire keríti őket hatalmába a félelem és a bizalmatlanság. A sikeres MI-alapú megoldások, legyen szó kereskedelmi vagy állami szolgáltatásokról, nem jöhetnek létre anélkül, hogy ne lenne meg a megfelelő szintű bizalom a használók, azaz a társadalom részéről. A bizalom pedig leginkább az ismereteken és az abban való hitben rejlik, hogy ezek a megoldások az életünket jobbá, és nem rosszabbá teszik.

Ezt a felismerést *Magyarország Mesterséges Intelligencia Stratégiája* is magáévá tette, hiszen célként tűzte ki széles társadalmi rétegek MI iránti érdeklődésének felkeltését, az emberek megismertetését az MI működésének főbb jellemzőivel, ami együtt kell, hogy járjon a kockázatok ismertetésével és tudatosításával is. Kutatásunk elsődleges motivációja az volt, hogy ezzel a komplex kihívással kapcsolatban legyenek nyilvánosan elérhető adatok a mesterséges intelligenciával kapcsolatos hazai társadalmi környezetről.

Ennek érdekében 2022 tavaszán egy 1250 fős országos reprezentatív, személyes megkérdezésen alapuló adatfelvétel során kerestük a választ arra, hogy milyen szintű a felnőtt magyar lakosság tapasztalata és ismerete a digitalizáció, ezen belül a mesterséges intelligencia fejlődésével és lehetőségeivel kapcsolatban, továbbá milyen általános várakozások kapcsolódnak az MI-n alapuló technológiákhoz.

2023 januárjában, a kutatássorozat második lépésében az információk „kínálati” oldala felé fordultunk, hiszen az új technológiákkal, így az MI-vel kapcsolatos vélemények és ismeretek alakulásában is meghatározó szerepe van a médiában megjelent információknak. A magyarországi kutatások eddig nem nagyon vizsgálták, hogy a hazai online média milyen formában számol be az MI alkalmazásáról, potenciális használati lehetőségeiről, és milyen látens és manifeszt üzeneteket közvetít az utópia-disztópia, illetve bizalom-bizalmatlanság dimenzióiban. Kutatásunk a 2020 és 2022 közötti időszakban megjelent híreket és közösségi média diskurzusokat elemzi és értelmezi.

### Résztvevő kutatók

Bokor Tamás  
Pálvölgyi Eszter  
Ságvári Bence

### Együttműködő partnerek

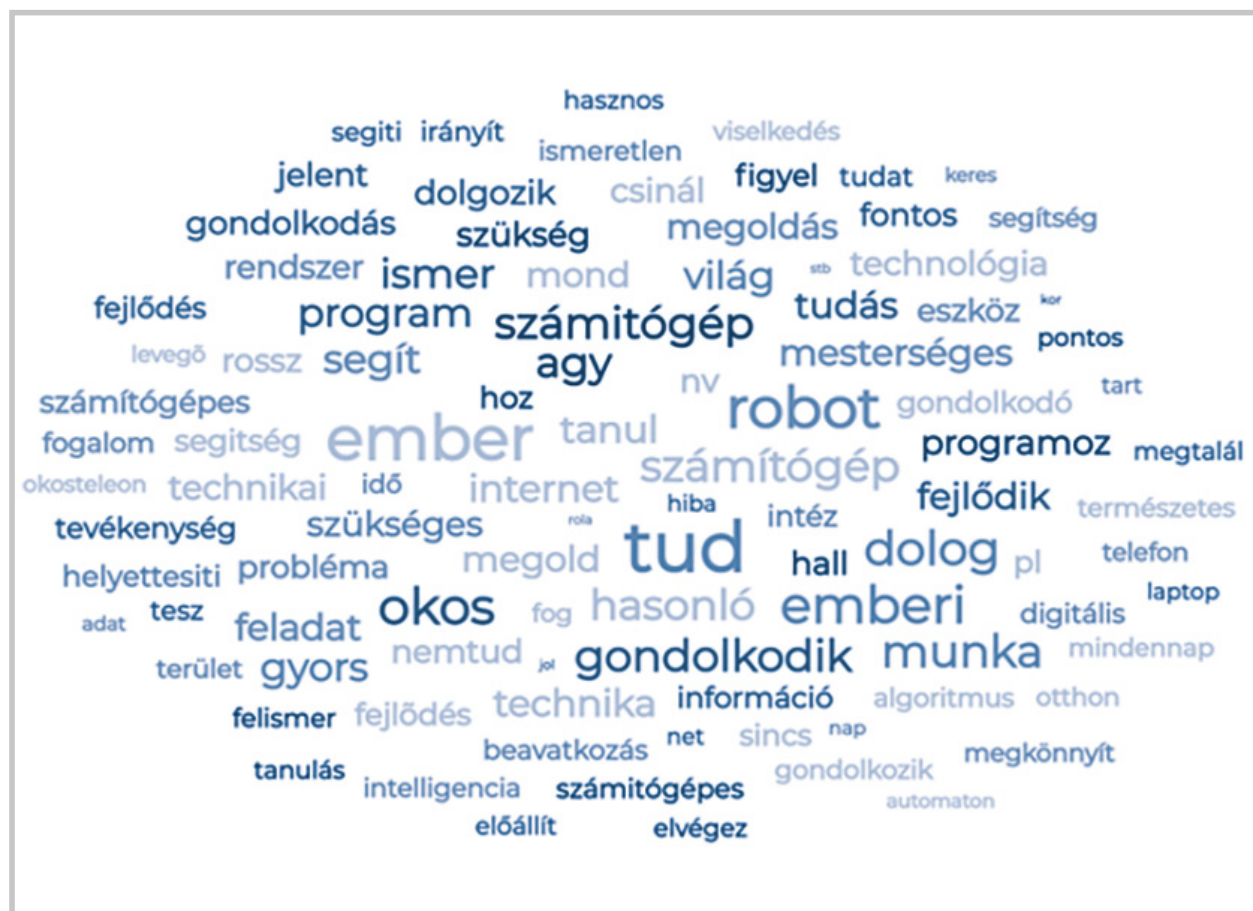
Budapesti Corvinus Egyetem  
ZRI Závecz Research Kft.

### Kutatási jelentés

Mi és az MI – Értékek, attitűdök, bizalmi kérdések a mesterséges intelligenciáról a magyar társadalomban



Mi és az MI Kutatási jelentés –  
Ságvári Bence: MI attitűdök és narratívák



A mesterséges intelligenciával kapcsolatosan említett szavak (Csak főnevek, igék és melléknevek)



# Politikai és jogi szövegbányászat mesterséges intelligencia segítségével

poltextLAB

Projektvezető: Sebők Miklós

<https://poltextlab.com/>



poltextLAB

A *Text Mining of Political and Legal Texts* (poltextLAB) projekt célja, hogy *Big Data* módszerekkel vizsgáljon magyar nyelvű és külföldi politikai és jogi dokumentum-gyűjteményeket. A kvalitatív adatforrások (szövegek, képek és videók) elemzésének hagyományos megközelítései jellemzően az adatok kézi feldolgozására építenek. Miközben a forrásanyag ismerete továbbra is nélkülözhetetlen bármilyen társadalomtudományi vizsgálatban, a kézi feldolgozás korlátai is nyilvánvalóak, elsősorban a kutatási eredmények megbízhatósága és érvényessége tekintetében. Az adatforrások széles skálája, nagy terjedelme (pl. egy ország minden elfogadott törvényének szövege) kivitelezhetetlenné teheti az emberi adatfeldolgozást. A kvantitatív szövegelemzési és szövegbányászati megközelítések a szöveges forrásokat alapul vevő társadalomtudományi *Big Data* projektek tekintetében így új módszertani standardot jelentenek.

A projekt során nagyméretű szövegkorpuszokat építünk, amelyeket elsődlegesen különböző, mesterséges intelligencián alapuló gépi tanító algoritmusok fejlesztésére és azok eredményességének tesztelésére használunk fel. Célunk magyar nyelvű szövegek elemzésére új módszertani megoldások kidolgozása, a meglévő algoritmusok eredményességének javítása, és új hibrid eljárások fejlesztése klasszifikációs feladatok megoldására, de egyben a módszerek kiterjesztése a nem magyar nyelvű korpuszok elemzésére is. Sikerral alkalmazzuk a különböző legkorszerűbb nagy nyelvi modelleket (pl. BERT) mind a magyar nyelvű, mind a többnyelvű klasszifikáció során, akár 20 osztályra is. A 2022-es év végén sikeresen bemutattuk legújabb innovációnkat, a *CAP BABEL MACHINE*-t, ami multilingual BERT-modell segítségével a *Comparative Agendas Project* (CAP) projekt szakpolitikai kódkönyvének fő témáit használja a szövegek szakpolitikai területeinek automatizált azonosításához. A [poltextlab.com/cap-machine](https://poltextlab.com/cap-machine) címen elérhető úrlapon keresztül a felhasználók

feltölthetik a kódolni kívánt fájlokat, majd az általunk kialakított folyamat segítségével rövid időn belül kézhez kaphatják a kész adatokat.

A projekt kiemelt célja a hasonlóan gondolkodó, szövegbányászati technikákat alkalmazó kutatókból álló hazai és nemzetközi hálózatot létrehozni. Évek óta fő szervezői vagyunk a nemzetközi *COMPTTEXT konferenciának*, amelynek célja, hogy lehetőséget biztosítson kutatásaikban a szövegbányászatot alkalmazó kutatók rendszeres találkozására, egymás eredményeinek megismerésére. Konferenciáinkon rangos nemzetközi egyetemokről és kutatóintézetekből évente közel 150 fő vesz részt. A projekt keretében évente 2–3 alkalommal tartunk szövegbányászati képzéseket társadalomtudósoknak. *Szövegbányászat és Mesterséges Intelligencia* képzési programunk kezdő és haladó szinten nyújt bevezetést a felügyelt és felügyelet nélküli gépi tanulási algoritmusok alkalmazásába. A program alapjául szolgáló, saját korpuszaink elemzésein alapuló *Szövegbányászat és mesterséges intelligencia R-ben* című tankönyvünk (szerzők: Sebők Miklós, Ring Orsolya, Máté Ákos) 2021-ben jelent meg. *Adatvizualizáció az R-ben* kurzusunk célja, hogy gyakorlatias és interaktív áttekintést nyújtson az R ggplot2 csomagját használó adatvizualizációról.

## Együttműködő partnerek

Jagiellonian University Kraków  
National University of Ireland, Galway  
Nemzeti Közzolgálati Egyetem  
Közigazgatástudományi Doktori Iskola  
Pécsi Tudományegyetem  
Microsoft AI Tudásközpont  
Reichman University  
Szegedi Tudományegyetem  
University of Cologne



## Publikációk

- Bolonyai Flóra, Sebők Miklós. *Kvantitatív szövegelemzés és szövegbányászat*. In: Jakab András, Sebők Miklós (szerk.) *Empirikus jogi tanulmányok*. Budapest: Osiris Kiadó, MTA Társadalomtudományi Kutatóközpont, 660 p., pp. 361–380., 20 p., 2020
- Gelányi Péter, Sebők Miklós, Ring Orsolya. *A topikmodellezés lehetőségei és korlátai egy törvénykorpusz példáján*. *Statistikai Szemle* 100: 8., pp 783–814., 2022
- Kiss Rebeka, Sebők Miklós. *Creating an Enhanced Infrastructure of Parliamentary Archives for Better Democratic Transparency and Legislative Research – Report on the OPTED forum in the European Parliament (Brussels, Belgium, 15 June 2022)*. *International Journal of Parliamentary Studies*, 2 (2), pp. 278–284, 2022
- Sebők Miklós, Boda Zsolt (szerk.). *Policy Agendas in Autocracy, and Hybrid Regimes*. London: Palgrave MacMillan, 2021
- Sebők Miklós, Gajdusczek György, Molnár Csaba (szerk.). *A magyar jogalkotás minősége: Elmélet, mérés, eredmények*. Budapest: Gondolat Kiadó, 400 p., 2020
- Sebők Miklós, Kacsuk Zoltán. *The Multiclass Classification of Newspaper Articles with Machine Learning: The Hybrid Binary Snowball Approach*. *Political Analysis* 29: 2, pp 236–249., 14 p., 2021
- Sebők Miklós, Kacsuk Zoltán, Máté Ákos. *The (real) need for a human touch Testing a human-machine hybrid topic classification workflow on a New York Times corpus*. *Quality and Quantity: International Journal of Methodology* 56, pp. 3621–3643., 23 p., 2022
- Sebők Miklós, Kozák Sándor. *From State Capture to „Pariah” Status? The Preference Attainment of the Hungarian Banking Association (2006–2014)*. *Business and Politics* 23: 2, pp 179–201., 2021
- Sebők Miklós, Kubik Bálint György, Molnár Csaba, Járay István, Székely Anna. *Measuring legislative stability – A new approach with data from Hungary*. *European Political Science* 21, pp. 491–521., 2022
- Sebők Miklós, M. Balázs Ágnes, Molnár Csaba. *Punctuated Equilibrium and Progressive Friction in Socialist Autocracy, Democracy and Hybrid Regimes*. *Journal of Public Policy* 42(2), pp. 247–269., 2022
- Sebők Miklós, Ring Orsolya, Máté Ákos. *Szövegbányászat és Mesterséges Intelligencia R-ben*. Budapest: Typotex Kiadó, 184 p., 2021

## Résztevő kutatók

Dinnyés Ágnes  
Járay István  
Gelányi Péter  
Kis György Márk  
Kiss Rebeka  
Kovács Ádám  
Kovács Viktor  
Kubik Bálint  
Lehoczki Richárd  
Máté Ákos  
Molnár Csaba  
Ring Orsolya  
Sebők Miklós  
Székely Anna  
Üveges István

## Repozitóriumok

- GitHub – [poltextlab/textreusel\\_ch\\_hun](#)  
GitHub – [poltextlab/text\\_mining\\_workshop](#)  
GitHub – [poltextlab/CLARIN\\_ParlaMint\\_HU](#)  
GitHub – [poltextlab/HunMineR: Companion package for the Hungarian text mining textbook](#)  
GitHub – [poltextlab/tankonyv: Szövegbányászat és mesterséges intelligencia R-ben](#)  
GitHub – [poltextlab/nyt\\_hybrid\\_classification\\_workflow: Replication material for Sebők, M., Kacsuk, Z., & Máté, Á. \(2021\). The \(real\) need for a human touch: testing a human-machine hybrid topic classification workflow on a New York Times corpus. Quality & Quantity, 1–23.](#)  
Multilingual comparable corpora of parliamentary debates ParlaMint 2.1

## Fontosabb konferenciák

- American Political Science Association (APSA) Political Methodology Specialist Group, Politics and Computational Social Science, Annual COMPTExT Conference, OPTED Data4Parliaments

## Elnyert nemzetközi pályázatok

- OPTED – Observatory for Political Texts in European Democracies (2020–2023) Horizon 2020 Grant Agreement no. 951832.

poltextLAB Tankönyv



# Octopus Research Tools (ORT)

Projektvezető: Ságvári Bence

<https://octopus-research.hu>

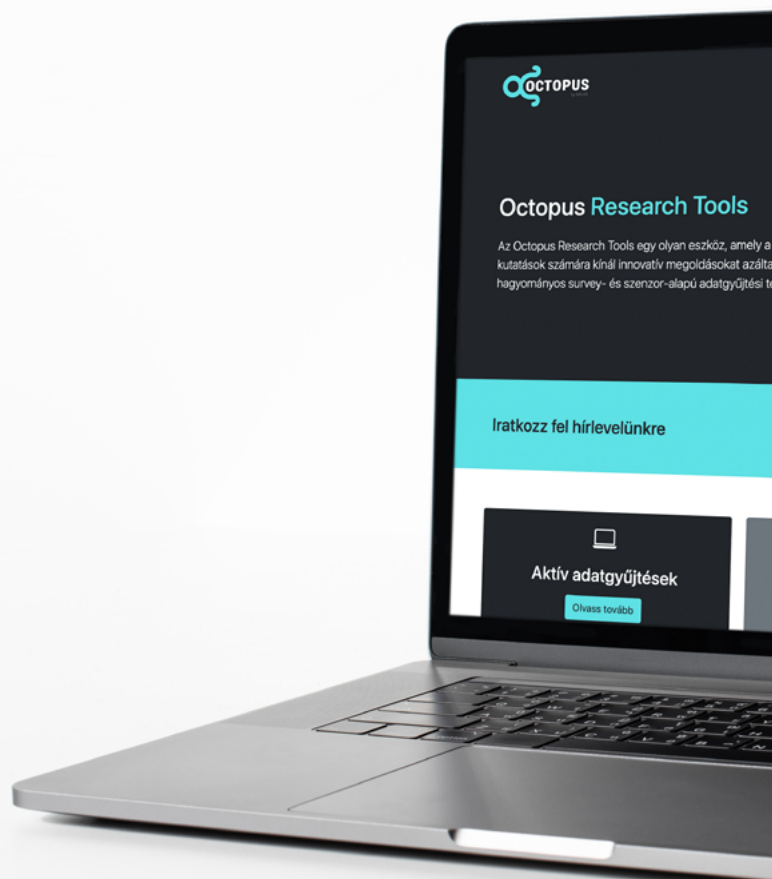
Az elmúlt években a személyes megkeresésen alapuló (face-to-face) tudományos adatgyűjtés világszerte válságba került. A módszertan megvalósítása egyre nehezebbé és költségesebbé válik. Ennek eredményeként egyre inkább előtérbe kerülnek a nem személyes megkeresésen alapuló adatgyűjtések, az emberi viselkedéssel kapcsolatos „digitális lábnyomokat” felhasználó kutatások. Ezek az adatok több forrásból származhatnak, melyek között megtalálható a weboldalak látogatottsága, a közösségi média aktivitás, az okostelefon alkalmazások használata és a készülékek különféle szenzoraiból származó mozgásra és egyéb fizikai körülményekre vonatkozó adatok. Az adatgyűjtés jellegét tekintve ezek nem-intruzív módszerek: a felhasználó részéről nem igényelnek folyamatos interakciót, ugyanakkor nagyfokú pontosságot és részletességet biztosítanak az emberek digitális életstílusáról.

Az Octopus projekt célja, hogy a szenzor- és készülékhasználati log alapú digitális viselkedésre vonatkozó adatgyűjtéseket és a multimédiás tartalmakat is megjeleníteni képes kérdőívek módszerét egy komplex, okostelefon-alapú szoftver ökoszisztéma keretében egyesítse. Ezzel egy nemzetközileg is innovatív kutatási eszközt hozunk létre, amely alkalmas a legkülönbézetesebb társadalmi kérdések és problémák tudományos és akár kereskedelmi célú vizsgálatára, az emberi viselkedés modellezésére és előrejelzésére, és ennek alapján a szakpolitikai döntéshozatal támogatására. A Mesterséges Intelligencia Nemzeti Laboratórium keretében 2021 őszén kezdődött el társadalomkutatók és informatikusok részvételével az *Octopus Research Tools* fejlesztése. Ezt megelőzően módszertani megalapozó kutatások zajlottak, illetve a rendszer funkcióinak specifikálására került sor. Ennek része volt a piacon elérhető szoftverek funkcióinak elemzése és újragondolása, továbbá saját, új funkciók kidolgozása.

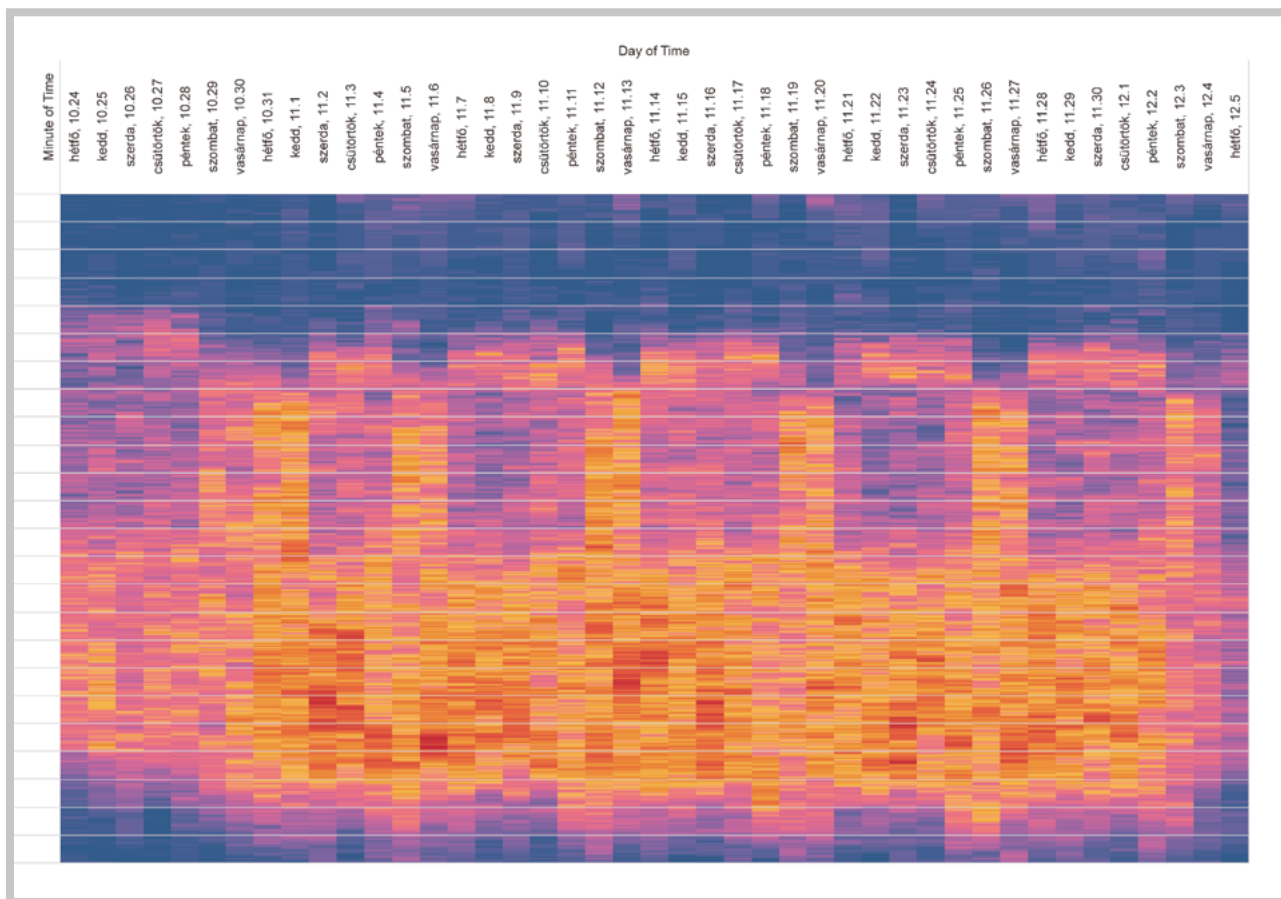
A legnagyobb magyarországi online kutatási panelen 2021-ben egy kísérleti kutatásra került sor, amely ún. vinyettás módszerrel vizsgálta az applikáció-alapú adatgyűjtéssel kapcsolatos válaszadói hajlandóság mintázatát, a kutatási ösztönzők szerepét, illetve arra kereste a választ, hogy melyek azok az adattípusok, amelyeket az emberek inkább, és amelyeket kevésbé vagy egyáltalán nem hajlandók megosztani tudományos, kereskedelmi és szakpolitikai célra. A kutatást elemző tanulmányban arra kerestük a választ, hogy a hagyományos kérdőíveken keresztül hozzáférhető személyes információkhoz, véleményekhez képest az emberek miként „árazzák” be saját digitális lábnyomaikat. Eredményeink azt mutatják, hogy az adatgyűjtés tartalma és kontextusa jelentősen megváltoztatja az emberek részvételi hajlandóságát, a különböző passzív szenzoros adatgyűjtési technikákhoz pedig más és más kockázatokat és megbízhatóságot társítanak. Ezeket a módszertani kutatásokat 2023-ban is folytatjuk, első körben egy 100 fős kísérleti pilot kutatás keretében. 2022 nyarára elkészült a szoftver általános keretrend-

szere, illetve az alapvető jogosultságok kezelése. 2023 januárjában már elérhető a komplex projektszerkesztő felület, a kérdőívszerkesztő és az adatgyűjtések időzítését szabályozó trigger és notification modul, továbbá a min-tavételezés. Az okostelefonos applikáció jelenleg Android készülékekre érhető el, 2023 során elkészül az iOS rendszert futtató iPhone készülékekre is. Az *Octopus Research Tools* alapműködésében nem használ MI-t, azonban több pontján is szorosan kapcsolódik ahhoz.

A különféle aktív és passzív adatgyűjtési modulok hozzájárulhatnak a minél hatékonyabb MI-alapú szolgáltatások létrehozásához. Komplex és transzparens módon lehet majd az emberektől adatokat gyűjteni, amelyek később különböző tanuló algoritmusok bemeneti adataiként is szolgálhatnak. A fejlesztés során arra törekszünk, hogy azoknál a funkcióknál, ahol ennek van jelentősége, megjelenjenek az MI-alapú funkciók.



Az Octopus Research Tools honlapja



Az okostelefon-használat intenzitásának hő térképe perc alapú bontásban 6 hetes időtartamban (aggregált felhasználói adatok)



## Résztevő kutatók

Gulyás Attila  
Kollányi Bence  
Koltai Júlia  
Ságvári Bence

## Publikációk

Egedy Tamás, Ságvári Bence. Urban geographical patterns of the relationship between mobile communication, social networks and economic development – the case of Hungary. *Hungarian Geographical Bulletin*. 70(2), 129–148., 2021

Ságvári Bence, Gulyás Attila, Koltai Júlia. Attitudes Towards Participation in a Passive Data Collection Experiment. *Sensors*. 21(18):6085., 2021

## Együttműködő partnerek

Nirprint Hungary Kft.  
NRC Marketingkutató és Tanácsadó Kft.  
SWT Informatika Kft  
TÁRKI Társadalomkutatási Intézet Zrt.





## A mesterséges intelligencia jogi kihívásai

Projektvezető: Mezei Kitti

fogyatékoságon, életkoron vagy szexuális irányultságon alapuló megkülönböztetésmentességet, a személyes adatok és a magánélet védelmét vagy a fogyasztóvédelmet, valamint a hatékony jogorvoslathoz és a tisztességes eljáráshoz való jogot. A kockázatok az MI-rendszerek általános tervezésének hibáiból vagy az adatok korrigálás nélküli felhasználásából adódhatnak. Ezen rendszerekbe biztonsági és tervezési mechanizmusokat kell integrálni az ellenőrizhető biztonság érdekében. A kutatás a tervezés útján történő szabályozás jogi és technikai feltételeit vizsgálja.

A projekt célja az EU mesterséges intelligenciáról szóló rendeletének a vizsgálata, mely jogszabály az EU-ban forgalomba hozott vagy használt valamennyi MI-rendszerre alkalmazandó horizontális minimumszabályokat határozza meg. Az MI rendelet egy kockázatalapú megközelítést alkalmaz, amelynek értelmében az MI-alkalmazásokat kockázati osztályokba sorolják, és a szükséges jogi beavatkozásnak a mértékét az adott kockázati szinthez igazítják. E célból megkülönböztetnek elfogadhatatlan, nagy kockázatú, közepes kockázatú és minimális kockázatú MI-alapú rendszereket.

Az MI-t alkalmazó rendszerek a Big Data elemzése és a benne rejlő lehetőségek kiaknázása révén javíthatják a döntéshozatali folyamatokat. Nagy mennyiségű ilyen információ különböző adatforrásokkal kombinálva lehetővé teheti az érintett azonosítását, és így az adatvédelmi jogszabályok megsértését eredményezheti. Mint minden informatikai rendszert, az MI-t is érhetik kiber-támadások, melyek gazdasági veszteséget és akár fizikai károkat okozhatnak (pl. MI-alapú robotok vagy MI-által vezérelt kritikus infrastruktúra rendszerek). Emellett az egyre szélesebb körben terjedő arcfelismerő technológia az *Általános Adatvédelmi Rendelet* (GDPR) értelmében „biometrikus adat” feldolgozásának minősül, és szigorú felhasználási feltételekhez kötött. Ezen adatok gyűjtése és felhasználása távoli azonosítási célokra, például arcfelismerésre nyilvános helyeken, az alapvető jogokat érintő kockázatokat hordoz magában.

A kutatás a digitális piac szabályozásának uniós belső piaci jogi aspektusaira és különösen az MI szerepére is összpontosít. A versenyjog, a fogyasztóvédelmi jog és a tisztességtelen kereskedelemmel kapcsolatos joganyag egyre bővül az MI hagyományos és digitális piacokon történő fokozott alkalmazása által okozott kihívásokkal kapcsolatban. A kutatás keretében sor kerül az uniós szabályozás dinamikájának vizsgálatára, különös tekintettel az MI szabályozására és annak a nemzeti jogalkotásra gyakorolt hatásaira.

Az alapkutatás mellett a projekt célja az ágazatspecifikus jogi kérdések felmérése, és a konzorciumi tagok számára szakmai tanácsadás. A projekt keretében tudományos és szakmai programokat szervezünk, elindítottuk a *Mesterséges intelligencia és jog* programsorozatot. 2022

A projekt a Mesterséges Intelligencia Nemzeti Laboratórium részeként támogatja a magyar mesterséges intelligencia stratégia célkitűzéseinek megvalósítását. A projekt célja a mesterséges intelligencia (MI) szabályozási környezetével kapcsolatos jelenlegi és jövőbeli problémák feltárása, a lehetséges irányok feltérképezése, a jó gyakorlatok azonosítása a különböző jogágakon belül az MI-vel kapcsolatban, összehasonlító módszerek alkalmazásával és kiemelten az EU jogi keretrendszerére fókuszálva. Ez alapján a kutatók konkrét és hatékony jogi ajánlásokat fogalmaznak meg az MI jogi és szabályozási környezetének előmozdítása érdekében.

Az MI-technológia sajátos jellemzői megnehezíthetik a hatályos nemzetközi, uniós vagy nemzeti jogszabályok betartásának ellenőrzését, azok hatékony érvényesítését. Az Európai Bizottság mesterséges intelligenciáról szóló fehér könyve szerint az MI alkalmazása az alapvető jogok további megsértéséhez vezethet, ideértve a véleménynyilvánítás szabadságához való jogot, a nemi, faji vagy etnikai származáson, valláson vagy meggyőződésen,

júniusában a Jogtudományi Intézet a magyarországi Francia Nagykövetséggel és Intézettel közösen rendezett nemzetközi konferenciát, mely platformot biztosított az MI jelenlegi és jövőbeli jogi szabályozásának és európai hatásainak megvitatásához. A konferencia előadásai írott formában 2023-ban jelennek meg. Ugyanezen évben jelenik meg a projektben részt vevő kutatók által írt kézikönyv az MI aktuális szabályozási kérdéseiről, valamint Mezei Kitty 2020-as monográfiájának átdolgozott és bővített kiadása is. A projekt során emellett több kutató rangos hazai és nemzetközi folyóiratokban publikált tanulmányokat.

## Publikációk

- Ambrus István. *Digitalizáció és büntetőjog*. (monográfia), Wolters Kluwer Hungary, Budapest, Magyarország, 328 p. 2021
- Ambrus István, Mezei Kitty. *The New Hungarian Legislation on Money Laundering and the Current Challenges of Cryptocurrencies*. *Danube*, vol.13, no.4, pp.256–268., 2022
- Chronowski Nóra, Kálmán Kinga, Szentgáli-Tóth Boldizsár. Régi keretek, új kihívások: a mesterséges intelligencia prudens bevonása a bírósági munkába és ennek hatása a tisztességes eljáráshoz való jogra. *Glossa Iuridica*, 8 (4). pp. 7–38. ISSN 2064-6887, 2022
- Czebe András. Fegyverek egyenlősége a digitális forradalom korában: a leplezett eszközök alkalmazásával összegyűjtött elektronikus adatok garanciális kérdései. *Kúriai Döntések Bírósági Határozatok: A Kúria lapja* 70: 1 pp. 141–147. 7 p., 2022
- Friedery Réka. Hardening Borders during the pandemic in the European Union: The Shape of the Schengen Borders Code and the EU Digital COVID Certificate. *Quarterly on Refugee Problems* 61: 1 p. 71., 2022
- Hoffman István, Karpiuk, Mirosław. E-administration in Polish and Hungarian Municipalities – a Comparative Analysis of the Regulatory Issues. *Lex Localis: Journal of Local Self-Government* 20: 3 pp. 617–640., 24 p., 2022
- Kálmán Kinga. The constitutional aspects of applying AI based decision-making in the judiciary through the lenses of the right to fair trial. In: Garayová Lilla (szerk.) *The Law of the Future – The Future of Law*. Bratislava, Szlovákia: Páneurópska vysoká škola, pp. 823–847., 2021
- Kálmán Kinga, Kiss Laura Olga, Szentgáli-Tóth Boldizsár. Mesterségesintelligencia-alapú szoftverek a világ bíróságain: gyakorlati tapasztalatok, perspektívák és kihívások. In: *Jogi technológiák*. Ludovika Press, Budapest. ISBN 9789635318193, 2022
- Kecskés Gábor. Az autonóm járművek jogi kérdéseinek nemzetközi kontextusa, különös tekintettel a környezetjogi vetületekre. *Állam- és jogtudomány* 61 (4). pp. 52–64., 2020
- Mezei Kitty. A kiberbűnözés aktuális kihívásai a büntetőjogban. L'Harmattan Kiadó. Budapest, Magyarország, 282 p. 2021
- Mezei Kitty. A modern technológiák kihívásai a büntetőjogban, különös tekintettel a kiberbűnözésre. *Állam- és jogtudomány* 61 (4). pp. 65–81., 2020
- Mezei Kitty. Diszkrimináció az algoritmusok korában. *Magyar Jog* 69 (6), pp. 331–338., 8 p., 2022
- Mezei Kitty, Bán-Forgács Nóra. *Discrimination in the Age of Algorithms*. In: *Human Rights as a Guarantee of Smart, Sustainable and Inclusive Growth*. Milton Friedman University; Alcide De Gasperi University of Euroregional Economy, Budapest; József, pp. 73–80., 2022
- Mezei Kitty, Krasznay Csaba. *Cybersecurity and Cybercrime in Hungary During the COVID-19 Pandemic*. In: Chałubińska-Jentkiewicz, Katarzyna; Hoffman István (szerk.) *The Role of Cybersecurity in the Public Sphere – The European Dimension*, Maribor, Szlovénia: Lex Localis, 234 p. pp. 191–207., 17 p., 2022
- Mezei Kitty, Szentgáli-Tóth Boldizsár. Az online platformok használatában rejlő veszélyek: a dezinformáció és a kibertámadások jogi kockázatai. In: *Demokrácia-dilemmák*. ELTE Eötvös József Kiadó, Budapest, pp. 241–261. ISBN 9789634894773; 9789634894766, 2022
- Rácz Lilla. A személy és a dolog fogalmának (lehetséges) változásai a mesterséges intelligencia és a kriptovaluták világában. *Állam- és jogtudomány* 61 (4). pp. 82–107., 2020
- Szentgáli-Tóth Boldizsár. A robotok jogi személyiségének koncepciója és annak lehetséges hatásai a demokráciára: Miért ne adjunk szavazati jogot a mesterséges intelligenciáknak. *Magyar Jog* 69 (6). pp. 321–330., 2022
- Szentgáli-Tóth Boldizsár. Digitalizáció az észtsországi bíróságokon: legenda és valóság. *Arsboni*, 10 (aug.). ISSN 2064-4655, 2022
- Szentgáli-Tóth Boldizsár. Egy dinamikus példa: mesterséges intelligencia Malajzia bíróságain. *Arsboni*, 10 (aug.). ISSN 2064-4655, 2022
- Szentgáli-Tóth Boldizsár. Robotic personhood and its potential impact to democracy. Should artificial intelligence be citizens and vested with right to vote? In: Garayová Lilla (szerk.) *The Law of the Future – The Future of Law*. Bratislava, Szlovákia: Páneurópska vysoká škola (2021) pp. 771–807., 2021
- Varju Márton. The Protection of Technology Sovereignty in the EU: Policy, Powers and the Legal Reality. *European Law Review* 47: 4 pp. 568–583., 16 p., 2022

## Résztvevő kutatók

Ambrus István  
Chronowski Nóra  
Czebe András  
Friedery Réka  
Győry Csaba  
Hoffman István  
Hoffmann Tamás  
Kecskés Gábor  
Kovács Szitkay Eszter  
Körtvélyesi Zsolt  
Mezei Kitty  
Pap András László  
Papp Mónika  
Polyák Gábor  
Rácz Lilla  
Szalay Klára  
Szentgáli-Tóth Boldizsár Artúr  
Szilágyi Emese  
Träger Anikó  
Varju Márton  
Zódi Zsolt



### Konferencia-előadások

- ICON-S Mundo Conference – The Future of Public Law, saját panel (AI and Law), online, 2021. július 6–9.
- Mezei Kitti. Diszkrimináció az algoritmusok korában. Európai Jogi Szaktanácsadó Hálózat éves továbbképzése, Budapest, 2021. november 22.
- Bán Marina, Szentgáli-Tóth Boldizsár. Artificial Intelligence and Legal Certainty: Promising Perspectives and Severe Risk. Factors The Impact of Digitalization on Constitutional Law, University of Copenhagen, Hybrid IACL Round Table, 2022. január 31. – február 1.
- Friedery Réka. Migration, pandemic and the prospects of Schengen. Global Migration: Trans-cultural, Trans-disciplinary and Intersectional Perspectives. Friedensau Adventist University (ThHF), online, 2022. február 9.
- Ambrus István, Mezei Kitti. Algorithmic decision-making in the criminal justice: risks and opportunities. The challenges of Artificial Intelligence for law in Europe: values, rights and regulation in the European legal space, Budapest, TK, 2022. június 9–10.
- Szentgáli-Tóth Boldizsár. The constitutionality of online court hearings. The challenges of Artificial Intelligence for law in Europe: values, rights and regulation in the European legal space, Budapest, TK, 2022. június 9–10.
- Mezei Kitti. The legal challenges of AI-enabled cyberthreats, Cyber Summit 2022, Varsó, 2022. június 14.
- Szentgáli-Tóth Boldizsár. Artificial Intelligence in criminal justice: recent experience from Malaysia. Legal Technology and its impact on the access to justice, society and the Legal Industry, Budapest, 2022. június 23.
- Mezei Kitti, Träger Anikó. The European and Chinese Approach to Artificial Intelligence: Risks and Opportunities. The 16th Annual Conference of the European China Law Studies Association, University of Copenhagen, Kopenhága, 2022. szeptember 21–23.

### Nemzetközi konferencia szervezése

The challenges of Artificial Intelligence for law in Europe: values, rights and regulation in the European legal space, Budapest, TK, 2022. június 9–10.

### Együttműködő partnerek

ELTE Állam- és Jogtudományi Kar  
Kúria  
Mollia Zrt.  
Nemzeti Közszoigálati Egyetem Eötvös József  
Kutatóközpont Kiberbiztonsági Kutatóintézet

### Nemzetközi előadássorozat

#### AI and Law programsorozat

#### Nemzetközi vendégkutatók online előadásai:

Molnár Tamás  
(European Union Agency for Fundamental Rights)

Joris Hulstijn (Tilburg University)

Béatrice Schütte (University of Helsinki),

Ljupcho Grozdanovski  
(NYU Law School/Université de Liège)

Adam Harkens (Birmingham University)

Pin Lean Lau (Brunell University, Law School)

Jakub Harasta (Masaryk University)

Rikke Jorgensen  
(The Danish Institute for Human Rights Copenhagen)

Wen Xiang (University of Copenhagen)

Reijer Passchier (Open University/Leiden University)



# A morális pánikgomb (MPG) modellezése hagyományos és NLP modellekkel

Projektvezető: Sik Endre

A kutatás célja a morális pánikgomb (MPG, egy, a kormány által létrehozott hosszú távú manipulációs technológia) működésének vizsgálata az online médiában és a Facebookon. A kutatás során egy már meglévő adatbázist (a SentiOne-től kinyert korpusz a 2019 és 2022 februárja között született, a Soros és/vagy a migráció kulcsszavakat tartalmazó cikkeket/posztokat tartalmazza) vizsgálunk részben hagyományos, részben természetes nyelvfeldolgozásra alapozott módszerekkel. Az adatbázison belül ezen túlmenően valamennyi részhalmazon belül a koronavírus-járvánnyal kapcsolatos cikkeket/posztokat magában foglaló alkorpuszokat is létrehozunk. A kutatás során az NLP különféle technikáinak alkalmazásával azt vizsgáljuk, hogy a kormányközeli és kormánytávoli szereplők, továbbá a három kulcsszó (Soros, migráció, koronavírus és ezek „szinonimái”) különféle kombinációiból előálló narratíva-proxyk hogyan függenek össze egymással, illetve az MPG létezésére utaló proxyk (speciális keretezés, jellegzetes fogalomhasználat, karaktergyilkosságok stb.) előfordulásának valószínűségével.

## Résztevő kutatók

Buda Jakab  
Farkas Eszter  
Molnár Anna  
Rakovics Zsófia  
Samoniek, Aleksandra  
Stefkovics Ádám

## Publikációk

Gerő Márton, Sik Endre. „Már nyomni sem kell” – a morális pánikgomb (MPG) és a 2022. évi választás. In *Az állandóság változása*, szerk.: Böcskei Balázs és Szabó Andrea, Gondolat/TK PTI, Budapest, 65–86 old., 2022

Kerényi Péter, Sik Endre. *Adalékok a morális pánikgomb legfrissebb megnyomattatásához. Mozgó Világ*, 11:15–27., 2022

## Konferencia-előadások

Sik Endre. The Symphony of the Moral Panic Button – Prelude and Four Movements, HWK Fellow Lecture, 2021

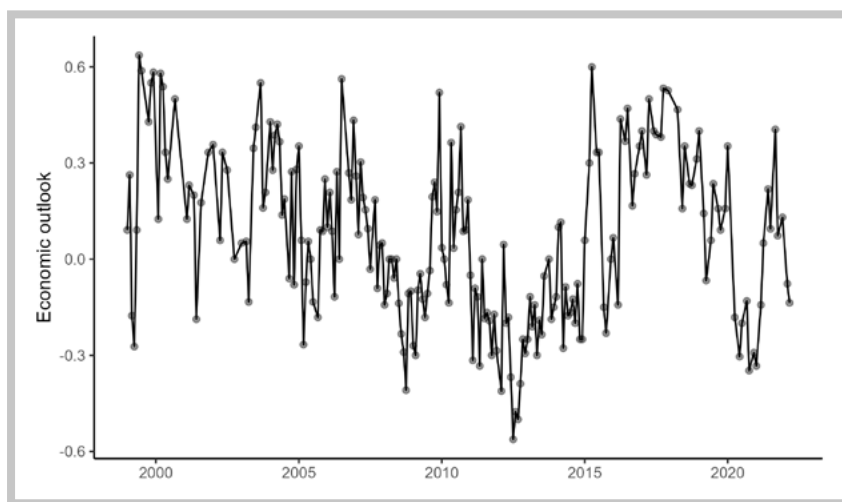


# Jegybanki kommunikáció mérése nagy nyelvi modellekkel

Projektvezető: Máté Ákos

A jegybanki kommunikáció kutatása egyre nagyobb hangsúlyt kap a közgazdaságtani és politikai gazdaságtani irodalomban, ám a jellemzően használt szótár alapú módszerek hatékonysága messze elmarad a felügyelt tanítási megoldásoktól. A projekt célja, hogy nagy mennyiségű „aranystandard” minőségű tanító adatok felhasználásával pontos képet adjon arról, hogy egyes szótárak milyen pontosak, illetve mekkora előrelépés várható, ha felügyelt tanulással történik a klasszifikáció. A felügyelt modellek részben a régóta használt *bag of words* alapú megoldások (pl.: *Naive Bayes*, *Random Forest* stb.), részben pedig a közelmúltban kidolgozott nagy nyelvi modellek, amelyek mély neurális hálókra alapulnak (pl.: BERT, RoBERTa).

A projekt eredményeként átfogó képet kapunk arról, hogy az egyes módszerek mennyire pontosan képesek megragadni a jegybanki kommunikáció látens dimenzióit. Ez azért különösen fontos, mert a pontos mérés a feltétele annak, hogy az ez alapján készült ökonometriai modellek eredménye megbízható legyen. Az eddigi előzetes eredmények alapján nagyságrendi eltérések vannak az egyes modellek között, amik jelentősen befolyásolhatják az ökonometriai modellek eredményét is. A végeredményben elkészülő finomhangolt nyelvi modell nyílt felhasználású lesz, ami a kutatáshoz tartozó adatrepozitóriumban lesz hozzáférhető.



Gazdasági szentiment az EKB elnöki beszédeiben

## Résztevő kutatók

Barczikay Tamás  
Máté Ákos

## Publikáció

Máté Ákos, Sebők Miklós, Barczikay Tamás. *The effect of central bank communication on sovereign bond yields: The case of Hungary*. *PLOS ONE*, 16 (2). pp. 1–28., 2021

## Konferencia-előadás

Máté Ákos, Barczikay Tamás. *European Central Bank communication during crises: ditching the boilerplate?*, 4th Annual COMTEXT Conference, Dublin, 2022. május 5–7.

## Repozitóriumok

GitHub – [poltextlab/central\\_bank\\_communication](https://github.com/poltextlab/central_bank_communication): Replication materials  
Jegybanki kommunikáció szentimentjét elemző script  
Monetáris szentiment elemzési R szoftvercsomag



# Magyar nyelvű benchmark adatbázisok gépi tanulás fejlesztésére

Projektvezetők: Kiss László és Molnár Csaba

A digitalizáció rohamos terjedésének, valamint a nyelvtechnológiai eszközök gyors fejlődésének köszönhetően az alkalmazott nyelvészeti kutatásokon belül az elmúlt időszakban kiemelt szerep jutott azon empirikus és adatvezérelt vizsgálati irányoknak, amelyek a társadalomtudományos kérdéseket valós szövegtestek (korpuszok) vizsgálatával kívánják megválaszolni. A részprojektben a CAP meglévő adatbázisait fejlesztettük tovább és benchmark adatbázisokat készítettünk el a meglévő adataink két irányban történő bővítésével. Egyrészt a már meglévő 2010 és 2022 közötti interpellációk, azonnali kérdések és napirend előtti felszólalások mellé bevontuk a (szóbeli) kérdéseket, másrészt az immár négy szövegtípusra vonatkozó adatainkat és korpuszainkat a 2002–2010 közötti két országgyűlési ciklussal is kibővítettük.

A nagyméretű kézzel vagy géppel annotált korpuszok alkalmas erőforrások lehetnek olyan korszerű módszerek alkalmazásához is, mint például gépi tanulási modellek tanítása. Ez utóbbiak közül a legfejlettebbek manapság szinte kivétel nélkül mesterséges neuronhálók alkalmazásán alapulnak, azonban hatékonysági okok miatt (az olyan nagy modellek használata, mint a BERT, sok esetben rendkívül számításgépes folyamat nem csak a betanítás, de akár a predikciós időben történő alkalmazás kapcsán is) a korábbi módszerek (logisztikus regresszió, szupertektor gép stb.) alkalmazása is mindmáig indokolható.

A kutatás során az elkészült korpuszokkal gépi tanítási kísérleteket végeztünk, hogy felmérhessük azok gyakorlati hasznát társadalomtudományi kérdések megválaszolásában. A módszerek széles spektrumát hasznosítottuk (pl. SVM, LSTM, BERT stb.) annak érdekében, hogy az egyes kutatási kérdésekhez a legalkalmasabb technológiát tudjuk kiválasztani. A kutatás során elkészült benchmark-adatbázis mellett az 1994–2022 közötti parlamenti felszólalások, törvényjavaslatok és törvények korpuszát, adatait tartalmazó Parlawspeech-adatbázis, a 2002–2014 közötti Magyar Nemzet-címlapok adatbázisa, valamint megjelent Molnár Csaba *If there is nothing else to say: the local content of interpellations* című tanulmánya a *Journal of Legislative Studies*-ban.

## Résztvevő kutatók

Kiss László  
Molnár Csaba  
Barczikay Tamás  
Kiss Rebeka  
Klein Adrienn  
Kovács Viktor  
Kubik Bálint György  
Pokornyi Zsanett  
Üveges István

## Publikáció

Molnár Csaba. *If there is nothing else to say: the local content of interpellations*. *The Journal of Legislative Studies*. Published online: 02 Oct 2022 pp. 1–23., Paper: Early Access, 23 p., 2022

## Konferencia-előadás

Boda Zsolt, Kiss László, Molnár Csaba. Nemzetközi komparatív adatbázisok elérhetősége a TK-ban – Bemutatkozik a TK Comparative Agendas Project-je. Szöveg. Gép. Társadalom, Budapest, ELTE Társadalomtudományi Kar, 2022. szeptember 20.

## Repozitórium

Hungarian PARLAWSPEECH dataset



Molnár Csaba



Kiss László



# Meddig érvényes egy klasszifikáció? Oltásellenes kommentek klasszifikációja

Projektvezető: Kmetty Zoltán

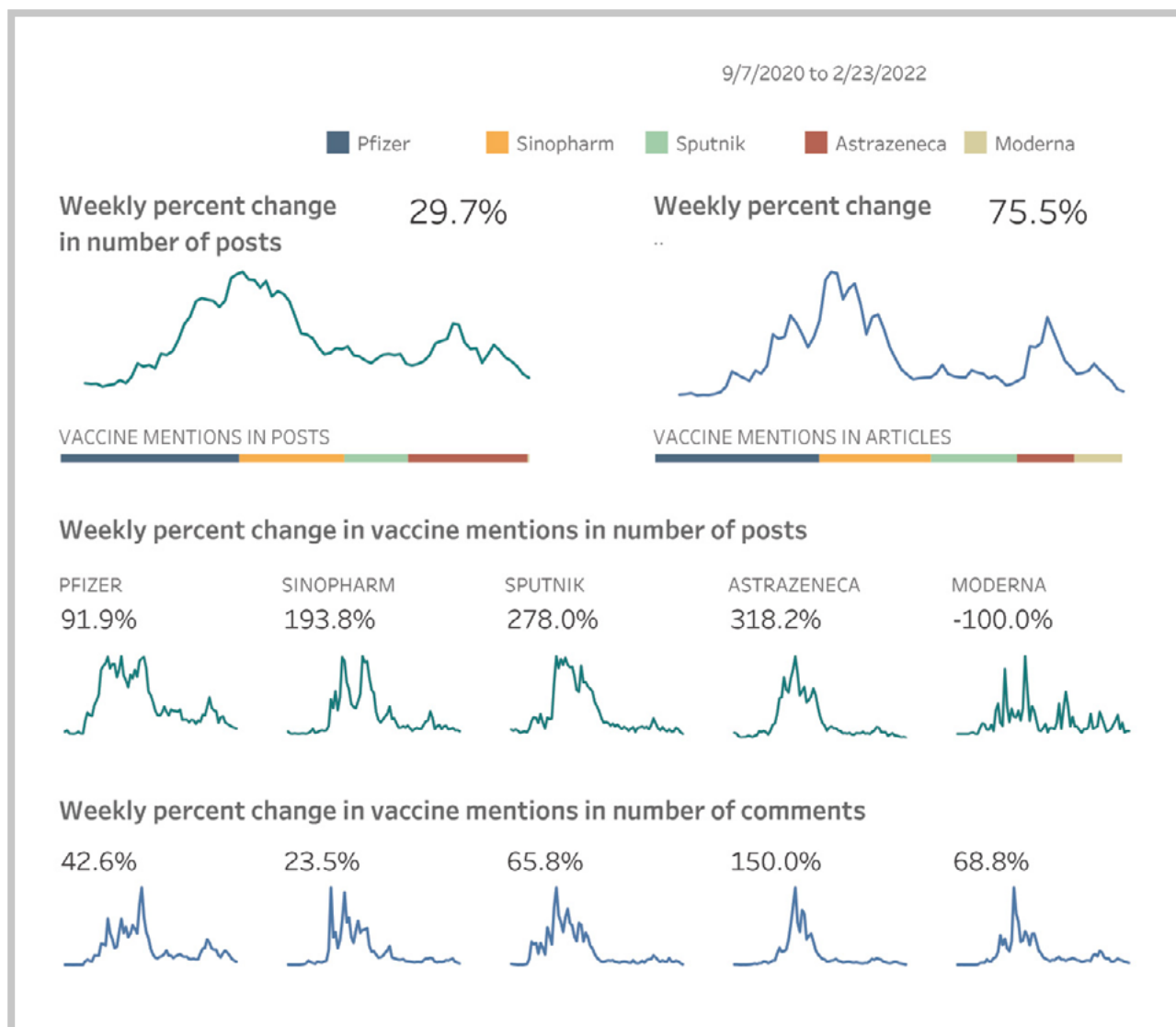
A COVID-dal kapcsolatos oltásellenességet vizsgáló projektünk tágabb célja az volt, hogy létrehozzunk egy olyan döntéshozó támogatási platformot, amely az online térben keletkező szöveges adatokat felhasználva képes az adott témában, valós időben képet adni a hazai lakosság véleményéről és reakciójáról. A klasszifikációs projektek nagy része időben már lezárult adathalmazok osztályozását végzi el. A dinamikusan növekvő digitalizáció viszont egyre több folyamatosan bővülő adatot tesz elérhetővé, ahol a klasszifikáció akár folyamatos igényként is felmerülhet. Ez a folyamatoság fontos jellemzője projektünknek is. Bizonyos klasszifikációs modellek képesek folyamatosan újranulni a bemenő adatok alapján, a tanuló adathalmazokra épülő modellek esetében azonban nehezebb az adaptáció. A modellek újratanításának magas a költsége, ezért fontos kérdés, hogy meddig használunk egy futó modellt, és mikor döntünk úgy, hogy új modellre van szükségünk.

Részprojektünkben a 2020–21-es időszakban kialakítottunk egy olyan keretrendszert, amely képes az

oltásokkal és oltásellenességgel kapcsolatos diskurzust monitorozni a magyar online térben. Ennek a keretrendszernek egy központi eleme az a BERT modell, ami a kommentek oltásellenesség szerinti klasszifikációját végzi el. A kiinduló modellt 2021 tavaszán készítettük el, 10 000 komment kézi klasszifikációjával. Mivel az adatgyűjtés folyamatosan zajlott, a diskurzus pedig időben változott, felmerült a modell működésének időtállósága. A kérdés megvizsgálásához 2021 végén készítettünk egy új annotációt a 2021 őszi kommentekre építve.

Ez az új annotáció lehetőséget teremt arra, hogy megvizsgáljuk, a különböző annotációk hogyan teljesítenek az egyes időszakokban, érdemes-e az annotált adatokat egyben kezelni, illetve hogyan működnek olyan speciális módszerek, mint a kommentek átírása más kategóriába. A 2022-es időszakban elsősorban azt vizsgáltuk, hogy mennyiben javítható címkézetlen adatokkal a komment klasszifikáció, milyen időszakból érdemes bevonni címkézetlen adatokat, és van-e értelme az annotálásnál a tartalmakat előszelektálni bizonyos szempontok mentén.





A MILAB projekt keretében kifejlesztett platform valós időben követte, hogy az online térben a COVID oltás kapcsán az egyes vakcinák mikor és mekkora arányban jelennek meg.

## Részvevő kutatók

Boros Krisztián  
Katona Eszter Rita  
Kmetty Zoltán  
Kollányi Bence  
Knap Árpád  
Vancsó Anna

## Publikáció

Vancsó Anna, Kmetty Zoltán. Dominant Christian narratives of solidarity during the COVID pandemic in Hungary. *Intersections. East European Journal of Society and Politics*. Vol. 7 No. 3, 2021, pp. 101–119

## Együttműködő partnerek

SentiOne  
Ynsight

## Workshop szervezése

Oltásellenesség három módszer keresztmetszetében, 2021. január 12., online meetup

## Konferencia-előadás

Kmetty Zoltán. A COVID-19 oltással kapcsolatos online diskurzus alakulásának vizsgálata szövegbányászati módszerekkel. VIII. Oktatás- és Kutatásmódszertani Szakmai Nap, Semmelweis Egyetem, 2021. február 3. Döntéstámogató platform demo bemutatója, MI Koalíció Szakmai Nap és Kiállítás, BME Q Épület, 2021. június 15–16.

Kmetty Zoltán. Mapping the network of anti-vaxxer and pro-vaxxer supporters. Sunbelt and NetSci Conference, online, 2021. július 6–11.

Kmetty Zoltán. „Russian vaccine will be good for our politicians; I want Pfizer” – Changing narratives of vaccination in Hungary. 7th International Conference on Computational Social Science IC2S2, online, 2021. július 27–31.

## Demo

MILAB COVID Dashboard

# Politikai szövegek véleményelemzése

Projektvezető: Ring Orsolya

A projekt fő célja különböző magyar nyelvű szövegek (online hírportálok, újságcikkek, politikusi beszédek, parlamenti felszólalások) elemzésére alkalmas szentiment- és emócióelemzési eljárások kidolgozása. Ezen eljárások célja az egyes szövegek tartalmából kinyerni azokat az információkat, amelyek értékelést fejeznek ki. Az elemzés különböző szinteken végezhető, részint attól függően, mi az elemzés alapegysége és meghatározzuk-e azt, hogy az érzelem mire irányul, vagy mi váltja ki azt. A szakirodalom különbséget tesz a pozitív-negatív-semleges skálán mozgó szentiment és a több kategóriával dolgozó érzelmelemzés között, melyek közül az utóbbi sokkal több információt nyújt az adott egység érzelmi töltetéről.

A feladat megvalósításához számos különböző módszerrel kísérletezünk. Egyrészt nagyméretű kézzel annotált szentiment- és emóciókorpuszokat hozunk létre, melyeket szótárak készítésére, gépi tanításra, valamint a gépi tanuló és a szótáralapú elemző algoritmusok eredményességének tesztelésére használunk fel. A projekt során *word embedding* segítségével, a szóbeágyazást

online hírportálok szövegein elvégezve készítettünk egy pozitív-negatív skálán mérő szentimentszótárat. Kidolgoztunk egy induktív érzelmekategorizáló rendszert, amely 12 különböző érzelmet különböztet meg a politikai szövegekben. Kategóriarendszerünk átváltható a nemzetközileg is használt emóciókategóriákra, így alkalmas az azokkal való összevetésre is.

Kettős vak kódolással elkészítettünk egy 5700 mondatból álló, mondatszinten annotált szentiment- és emóciókorpuszt, valamint egy tagmondat szinten annotált parlamenti beszédekből épített szentiment- és emóciókorpuszt (HunEmPoli), ahol az azonosított 39840 érzelmet azok argumentumaival is összekötöttük. Mindkét korpuszunk annotálása szigorú minőségbiztosítás mellett, nagy kódolók közötti egyetértéssel történt. A projekt jelenlegi szakaszában a HunEmPoli korpuszból képzett tanítóadatok segítségével a huBERT-modell finomhangolásával egy szentiment- és emócióelemző modell tanításán dolgozunk. Parlamenti beszédkorpuszunk vizualizálására egy internetes felületet is létrehoztunk:

<https://napirendek.hu/erzelmek/>.

## Résztvevő kutatók

Guba Csenge  
Ring Orsolya  
Szabó Martina Katalin  
Váradi Bendegúz  
Vincze Veronika

## Publikációk

Ring Orsolya, Vincze Veronika, Guba Csenge, Üveges István. *HunEmPoli: magyar nyelvű, részletesen annotált emóciókorpusz*. In: Berend Gábor, Gosztolya Gábor, Vincze Veronika (szerk.) *XIX. Magyar Számítógépes Nyelvészeti Konferencia*, Szegedi Tudományegyetem, Informatikai Intézet, 2023

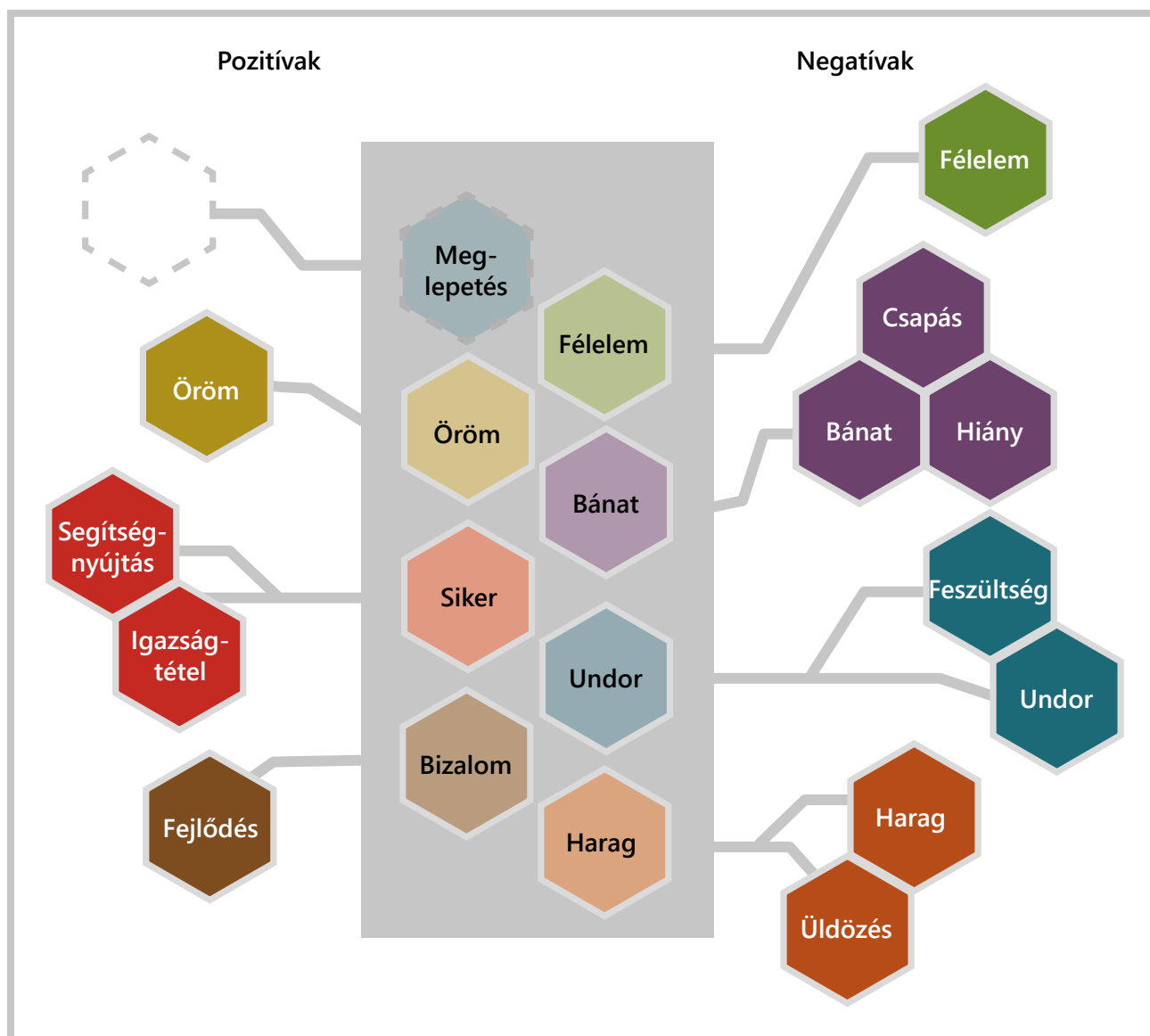
Szabó Martina Katalin, Vincze Veronika, Ring Orsolya, Guba Csenge. Nagyot mondó képviselők? Fokozás a politikai kommunikációban. In: Berend, Gábor; Gosztolya, Gábor; Vincze, Veronika (szerk.) *XVIII. Magyar Számítógépes Nyelvészeti Konferencia*, Szegedi Tudományegyetem, Informatikai Intézet, 2022

Üveges István, Vincze Veronika, Ring Orsolya, Guba Csenge. *Aspect-based emotion analysis of Hungarian parliamentary speeches*. In: *Proceedings of the 2nd Workshop on Computational Linguistics for Political Text Analysis*, Potsdam, 2022

Ring Orsolya (Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya et al.) *The ParlaMint corpora of parliamentary proceedings. Language Resources and Evaluation*, 2022







A HunEmPoli kopuszban használt emóciókategóriák megfeleltetése a Plutchik-féle kategóriarendszernek

## Konferencia-előadások

- Ring Orsolya, Guba Csenge, Vincze Veronika, Üveges István. HunEmPoli: magyar nyelvű, részletesen annotált emóciókorpusz, XIX. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, 2023. január 26-27.
- Szabó Martina Katalin, Vincze Veronika, Ring Orsolya, Guba Csenge. Nagyot mondó képviselők? Fokozás a politikai kommunikációban. XVIII. Magyar Számítógépes Nyelvészeti Konferencia, online, 2022. január 27-28.
- Üveges István, Vincze Veronika, Ring Orsolya, Guba Csenge. Aspect-based emotion analysis of Hungarian parliamentary speeches. KONVENS 2022, 2nd Workshop on Computational Linguistics for Political Text Analysis, Potsdam, 2022. szeptember 12-15.

## Együttműködő partnerek

- Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék, SmartLAB
- Charles University, Prága
- Kempelen Institute, Pozsony
- Montana Tudásmenedzsment Kft.
- Vistula University, Varsó

## Repozitórium

- A novel cost-efficient use of BERT embeddings in 8-way emotion classification on a Hungarian media corpus – Github
- Aspect based emotion analysis of Hungarian parliamentary speeches – Github
- HunEmPoli corpus – Github
- Possibilities and limitations of a lexicon-based sentiment analysis of Hungarian political news – Github

# Szöveges kutatási források tematikus feltárása a TK adatrepozitóriumában

Projektvezető: Gárdos Judit

Pilot projektünk célja a 20. Század Hangja és a TK Kutatási Dokumentációs Központ (KDK) digitális társadalomtudományos archívumokban őrzött interjúk gyűjteményekből vett mintán végzett különböző gépi szövegelemzési technikák tesztelésével az archívumokon belüli kereshetőség javítása volt. A projekt a KDK és a SZTAKI Elosztott Rendszerek Osztályának együttműködésében valósult meg. A legmegfelelőbb technika kiválasztása, alkalmazása és validálása után az eredményeket egy repozitóriumi kereső bétaverziójába integráltuk. Végeredményben az egyes interjúkhoz gépileg hozzárendeltünk olyan metaadatokat, amelyek információt adnak a szövegek (interjúk, interjúrészletek) tartalmáról, az egymással (akár több gyűjteményen át) összefüggő, adott kutatási kérdések szempontjából releváns szövegek, szövegrészek helyéről a kutatók számára.

Az interjúk szövegeihez előbb manuálisan, majd gépi segítséggel tárgyszavakat, tárgymutatót generáltunk, melyek megfelelőségét kutatóink ellenőrizték. Az eredmények validálása a kutatók által utólagosan nem ellenőrzött, kizárólag gépileg elemzett dokumentumok esetében is jól használható tárgyszavazást, tárgymutató-generálást eredményez. Az interjúkhoz társított tárgyszavak vagy címkék nem egyszerűen a szövegben megtalálható kulcsszavak vagy azok szinonimái, hanem egy nemzetközi társadalomtudományos tezaurusz, az ELSST használatával készült fogalmi háló elemei, melyek révén feltárhatók a szövegekben rejlő szociológiai jelenségek. Emellett lépéseket tettünk a NER (*Named Entity Recognition*) irányába is. A szövegekben azonosítottunk névelemeket és időmegjelöléseket, majd wikifikáltuk, vagyis Wikidata tudásgráfhoz, illetve Geonames, VIAF, PIM és más névterekhez kapcsoltuk őket.

A jobb kereshetőség érdekében a gépi feldolgozással kapott absztrakt tárgyszavakat és névelemeket több új metaadatmezőben társítjuk a dokumentumokkal. Ezáltal a meglévő dokumentumok új kutatások számára nyílnak meg. A tárgyszavakat angolra is lefordítjuk, így archívumaink külföldi kutatók számára is kereshetővé válnak. Ezáltal a nyelvi elszigeteltség miatt eddig hozzáférhetetlen hazai forrásokat teszünk láthatóvá és elérhetővé a nemzetközi kutatói közösség számára. A gépi feldolgozás eredményeit vizuálisan is bemutatjuk: egyrészt a névelemek kiemelésével és szócikkekhez kapcsolásával, másrészt feltárva és megjelenítve az egyes témák, tárgyszavak gyakoriságát és a közöttük lévő kapcsolatokat.

A projekt kapcsán csatlakoztunk a CESSDA (*Consortium of European Social Science Data Archives*) munkájához. Az együttműködés során a Nyelvtudományi Kutatóközponttal (NYTK) való kooperációban elkészült az ELSST több mint 3300 kifejezést tartalmazó, angol nyelvű társadalomtudományos tezaurusz magyar fordítása, amely



2022. szeptembere óta online is elérhető. Projektünkben továbbá együttműködtünk a BME-vel, hogy a BEAST (BEA – mint BESzélnyelvi Adatbázis – Speech Transcriber) magyar beszédleírató szoftver hatásfokát növeljük. A BEAST OTKA és MILAB támogatással, a NYTK és a BME kooperációjával készülő, nyílt forráskódú, kutatási célokra szabadon használható rendszer, a SpeechBrain kódjára épül, és a legmodernebb transzformer neurális struktúrákat használja. A szociológiai források iránt érdeklődő kutatók számára az interjúk dokumentumokat feltáró munkánk eredménye a TK KDK repozitóriumi számára készülő közös online keresőfelületen lesz látható.

## Résztvevő kutatók

Annus Szabolcs, Antal Emese, Egyed-Gergely Júlia, Filep Georgina, Gárdos Judit, Havadi Gergő, Horváth Anna, Jakab Miklós, Lipp Veronika, Matyasovszky Németh Márton, Meiszerics Enikő, Neményi Mária, P. Tóth Tamás, Sass Bálint, Szöllősi Melinda, Vajda Róza

## Publikáció

Egyed-Gergely Júlia, Vajda Róza, Gárdos Judit, Horváth Anna, Meiszerics Enikő, Micsik András, Martin Dániel, Marx Attila, Pataki Balázs, Siket Melinda. *Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok*. In: Tick, József; Kokas, Károly; Holl, András (szerk.) *Valós térben - az online térért: Networkshop 31: országos konferencia*. 2022. április 20–22. Debreceni Egyetem. Budapest, Magyarország, HUNGARNET Egyesület, MTA Könyvtár és Információs Központ, 364 p. pp. 161–169., 2022

## Konferencia-előadások:

FAIRsFAIR (EOSC alprojekt) Final Event. The National Perspective, online kerekasztal-beszélgetés, Gárdos Judit, 2022. január 26.

Egyed-Gergely Júlia, Micsik András, Vajda Róza: Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok - előadás, Networkshop 31: országos konferencia, 2022. április 20–22.

## Társadalmi-kognitív hatások a jogi döntéshozatalban: formális modellezés

Projektvezető: Péli Gábor

Az exploratív kutatás tárgya, a jogi fogalmak elmosódottságának alakulása egy olyan terület, mely eddig ellenállt a formális modellezési kísérleteknek. A fogalmi többértelműség mérhetővé tételével, alakulási dinamikájuk megismerésével az egységes, kiszámítható jogi működés döntéshozatal kialakítására irányuló eszköztár bővül. A jogi fogalmak jelentésváltozatait a szemantikai tér feletti sűrűségfüggvényekkel modellezzük, a jogi többértelműség dinamikáját pedig e függvények alakjának és térbeli távolságainak változásaival. Hasonló formális módszertani megközelítéssel több, velünk kapcsolatban álló külföldi kutatóhelyen is foglalkoznak, például a többértelműséget különféle művészeti területek fogalmi besorolási mintázatainak elemezve. E külföldi kutatások szintén a terület-specifikusan kialakított elméleti keretek tesztelési szakaszában vannak. Ezért kutatásunk egyben egy új, formálódó konzorcium hazai lábának kiépítését célozza, melynek munkájához mi a jogi fogalmi dinamika elemzésével járulnánk hozzá. A formális modellezést és az ezt követő kézi adatgyűjtésen alapuló empirikus anyagokat Deep Learning kontextusbeli tesztelések követik.

A kutatás indító szakaszában jogi alapfogalmak jelentésváltozatait vizsgáltuk, azok jellemző kognitív dimenzióit azonosítottuk. Ehhez az Alkotmánybíróság által végzett részletes jogi fogalom értékeléseket elemeztük, konkrét leíró dimenziókat keresve. Az első szakasz kognitív térbe való leképezésre vonatkozó tapasztalatait a hazai büntetőjogi ítéletek másodfokú, illetve Kúria-szintű ítéleteire alkalmaztuk (pilot study). A cél a fogalmak operacionalizálása, kognitív térbeli dimenzióinak tisztá-

zása volt. A kutatás feltáró szakaszának eredményeiről és a kiindulásul szolgáló formális modell ennek fényében történő adaptálásáról szóló előadásunk egy berlini nemzetközi konferencián hangzott el 2022 júniusában.

A fentieken túl a vázolt kognitív keretbe helyezett fogalmak jelentésváltozatainak előfordulási gyakoriságairól gyűjtöttünk adatokat a büntetőjogi adatbázisban, kézi munkával. Az előfordulási valószínűségek ismerete lehetővé teszi a jogi fogalmak kognitív távolságaira vonatkozó becsléseket (Kullback-Leibler divergencia alapján mért relatív entrópia számítás). Az empirikus munkát két angol nyelvű cikk leadásával és egy magyar cikk jogtudományi lapba való leadásával adjuk közre. Egyidejűleg a kézi kódolás folytatása mellett további kutatási anyagokat készítünk. A gépi tanuláson alapuló kódolás előkészületeinek megkezdése során támaszkodunk az említett berlini konferencia résztvevőinek párhuzamos kutatásaira, melyek ugyanazt a formális kognitív modellt adaptálják – de nem jogi fogalmi környezetekre.

### Résztvevő kutatók

Közantal Viktória  
Lőrincz Viktor  
Péli Gábor  
Pólos László

### Konferencia-előadás

Pólos László. Perception, Communication, and Inferencing. 2022 Nagymaros Conference, ESMT Berlin, 2022. június 19–22.





## Mesterséges intelligencia és drónhasználat a mezőgazdaságban

Projektvezető: Kovács Imre

A precíziós gazdálkodás az agrárinnováció lényeges eleme, melynek a legfontosabb jellemzője, hogy a gazdálkodás minden elemében kiemelt szerepe van a térinformatikai és távérzékelési megoldásoknak, a pontos méréseknek és a számítógépes vezérlésnek. A precíziós mezőgazdaság az információs technológia, a mesterséges intelligencia mezőgazdasági megjelenéseknek legfontosabb példája, olyan mezőgazdasági forradalomnak tekinthető, amelynek három meghatározója a gazdasági eredményesség és hatékonyság fokozása környezetbarát módon és társadalmi előnyök elérése érdekében. A precíziós gazdálkodás emellett társadalmi célokat is követ: elősegíti az élelmiszer-minőség és -biztonság növelését, a csökkenő környezetterhelés révén hozzájárul a klímaváltozással és a természeti erőforrásokkal történő hatékonyabb gazdálkodáshoz. Az MI-n alapuló technológia alkalmazása ezzel párhuzamosan környezet- és talajkímélő, növelve a jövedelmezőséget.

A két kutatási téma a precíziós gazdálkodás adaptálásának (például precíziós talajminőség és hozamtérképezés) társadalmi feltételei és a terjedésének hálózatelemzése. A társadalmi feltételek kutatásának az alapkérdése, hogy a gazdák egyes csoportjai – innovátorok (feltalálók), korai alkalmazók, a korai többség, a késői többség és a lemaradók – esetében milyen indikátorok szerint lehet meghatározni az egyes adaptációs csoportok precíziós technológiák alkalmazásra való nyitottságának jellemzőit és milyen tényezők befolyásolhatják a gazdák motivációit a mesterséges intelligencia felhasználásán alapuló precíziós gazdálkodásra történő átállásban. A második kutatási téma a precíziós gazdálkodás elterjedésének hálózati összetevőit elemzi. Ennek része a módszerre vonatkozó információk terjedésének, hálózati meghatározottságának a kutatása, illetve a részleges vagy teljes átváltás terjedésének hálózati elemzése.

2021 december és 2022 január vége között 200 precíziósan gazdálkodó növénytermesztővel készült kérdőíves adatfelvétel. A mintába került gazdaságok regionálisan és méretük szerint is reprezentálják a növénytermesztő magyar farmokat. A kérdőív részletes információkat nyert ki az MI különböző formáinak az alkalmazásáról. Ezek közül különösen a drónhasználat motivációjáról és gyakorlati vonatkozásairól kérdeztünk: ez az egyik legkésőbb megjelenő módja a mezőgazdaságban az MI használatának, viszont terjedése gyorsabb, mint más kiték. A kérdőíves felvételt 30 magnós interjú egészíti ki, amelyek precíziós gazdákkal készültek. A kutatás első elemzése a drónhasználat motivációival foglalkozik. A kutatásban alkalmazott transzelméleti modell (ordinális logit regressziós modell) használata, valamint a kérdőívünk szerkezete és kérdései alapján eredményeink alkalmasak egy német kutatással történő összehasonlításra. A német kérdések jelentős részét a mi kérdőívünkben is alkalmaztuk. A matematikai, statisztikai elemzés eredményeinek értelmezéséhez a kvalitatív interjúk információit is használjuk. A kutatás eredményei alkalmasak a közvetlen gyakorlati felhasználásra, rámutatnak a drónhasználat terjedését elősegítő tényezők jellegére és együttes hatására.

### Résztvevő kutatók

Bai Attila  
Balogh Péter  
Czibere Ibolya  
Gabnai Zoltán  
Kovács Imre  
Loncsák Noémi  
Megyesi Boldizsár  
Nemes-Zámbó Gabriella

### Publikációk

Bai Attila, Kovács Imre, Czibere Ibolya, Megyesi Boldizsár, Balogh Péter. [Examining the Adoption of Drones and Categorisation of Precision Elements among Hungarian Precision Farmers Using a Trans-Theoretical Model](#). *Drones* 6 (8), 200. 2022

Balogh Péter, Bai Attila, Czibere Ibolya, Kovács Imre, Fodor László, Bujdos Ágnes, Sulyok Dénes, Gabnai Zoltán, Birkner Zoltán. [Economic and Social Barriers of Precision Farming in Hungary](#). *Agronomy*. 11(6):1112, 2021

Kovács Imre, Megyesi Boldizsár, Bai Attila, Balogh Péter. [Sustainability and agricultural regeneration in the Hungarian agriculture](#). *Sustainability* 14: 2 Paper: 969, 14 p., 2022

### Együttműködő partnerek

Debreceni Egyetem  
Kynetec Hungary Kft

## Munka 5.O

# A technológia és a munka a mesterséges intelligencia korában

Projektvezető: Illéssy Miklós

Az elmúlt években intenzívebbé váltak a digitalizációval és a mesterséges intelligenciával, valamint ezek társadalmi hatásaival foglalkozó kutatások és azok eredményeit kísérő viták. A fejlett, ún. smart technológiák használatától remélt jelentős termelékenységnövekedés egyelőre várat magára. A lehetséges magyarázó tényezők közül az egyik legfontosabb, hogy a technológiai és a társadalmi rendszerek együttes optimalizációja elmaradt, utóbbi nem tud lépést tartani a technológiai fejlődéssel. A gondolat nem új, már a második világháború alatt foglalkoztak vele a londoni Tavistock Intézetben. Az ezekből a kutatásokból született elmélet szerint kiemelt fontosságú, hogy a vállalati gyakorlatban összehangolják a technológiát, az embereket és a munkaszervezetet. Ezt nevezik a társadalmi-technológiai rendszerek dizájnjának (*sociotechnical system design*).

Kutatásunkban arra vállalkozunk, hogy nagymintás európai kérdőíves felmérések (*European Company Survey, European Working Conditions Survey*) adatbázisain végezzünk másodelemzéseket. Az elemzések a következő kérdésekre keresik a választ: hol tartanak a magyar vállalatok a digitális áttérés gyakorlatában, mennyire elterjedtek az úgynevezett tanuló szervezetek Magyarországon, a magyar vállalatok milyen mértékben támaszkodnak munkavállalói tudására, milyen eszközökkel igyekeznek azokat mozgósítani, milyen összefüggés fedezhető fel a digitális áttérés, a munkavállalói részvétel és a vállalati innovációs teljesítmény között.

E kérdésekre adott válaszok segítségével a digitális technológiák munkaszervezetre gyakorolt hatásait, valamint az olyan társadalmi-szervezeti (ún. puha) tudások fontosságát szeretnénk a munkavégzés praxisában (menedzsment gyakorlatában) vizsgálni, mint a kommunikációs és kooperációs készség, az egyéni és kollektív tudások fejlesztése, a komplex feladatok megoldásának képessége, vagy a tárgyalási-meggyőzői képesség. Az európai adatbázisok elemzésével mindezt a nemzetközi összehasonlítás perspektívájából tehetjük majd meg, amelynek révén lehetővé válik nem csak az európai átlaggal való összevetés, de a hozzánk hasonló fejlődési utat bejáró visegrádi országcsoporttal való összehasonlítás is.

### Nemzetközi online workshop szervezése

Knowledge Infrastructure in the Platform Economy, online, 2022. április 21. Az NKE-vel közös konferencia, amerikai, portugál, szerb, német, finn, ausztrál indiai résztvevőkkel, a mesterséges intelligencia és algoritmus menedzsment munkahelyi felhasználásának módjairól és hatásairól.

### Résztvevő kutatók

Illéssy Miklós  
Makó Csaba  
Nostratabadi, Saeed  
Pap József

### Publikációk

Illéssy Miklós, Huszár Ákos. Technológiai fejlődés és munkaerőpiac: automatizációs kitettség Magyarországon. *Statisztikai Szemle* 100: 2 pp. 137–161., 25 p., 2022

Illéssy Miklós, Huszár Ákos, Makó Csaba. Technological development and the labour market: how susceptible are jobs to automation in Hungary? *Societies* 11(3):93, 2021

Makó Csaba, Illéssy Miklós, Pap József. Munkavégzés a platformalapú gazdaságban – A foglalkoztatás egy lehetséges modellje? *Közgazdasági Szemle* 67 (11). pp. 1112–1129, 2020

Makó Csaba, Illéssy Miklós, Nostratabadi, Saeed. Is it Possible to Improve Simultaneously Firm Performance and Workplace Well-Being? Small and Medium-Sized Enterprises in European Comparison. In: Dallago, B., & Casagrande, S. (Eds.). *The Routledge Handbook of Comparative Economic Systems* (1st ed.). Routledge, 2022

Pap József, Makó Csaba, Illéssy Miklós, Dedaj, Zef, Ardabili, Sina, Török Bernát, Mosavi, Amir. Correlation Analysis of Factors Affecting Firm Performance and Employees Wellbeing: Application of Advanced Machine Learning Analysis. *Algorithms* 15(9), 300, 17 p., 2022

### Konferencia-előadások

Makó Csaba, Illéssy Miklós. Automation, Creativity, and the Future of Work in Europe: A Comparison between the Old and New Member States with a Special Focus on Hungary. International Conference on „Technology, Market and Complexity”, Society of Open Innovation & Riga Technical University, 2021. július 12–15.

Makó Csaba, Illéssy Miklós, Nostratabadi, Saeed. Is it Possible to Improve Simultaneously Firms' Performance and Well-being in Work? (A Structural Equation Modelling-based analysis of the European Company Survey 2019), The new global challenges and comparative economics, University of Trento, 2021. október 18–20.





## Oksági gépi tanulás: az alapelvektől a datasheet-ig

Projektvezető: Janky Béla

Az utóbbi évtizedben a társadalomtudományi statisztika két területén zajlottak a statisztika uralkodó paradigmáját megkérdőjelező, jelentős hatást gyakorló kutatások. Egyrészt a strukturális oksági modellezés mutat új utat a társadalomkutatási gyakorlatban, másrészt a mérnöki területeken kidolgozott gépi tanulási eljárások társadalomtudományi alkalmazása forradalmasítja az ökonometriát. A két terület látszólag egymással ellentétes fejlődési utakat jelöl ki a társadalmi adatok elemzői számára. A strukturális oksági modellek a jelenlegi gyakorlatnál jóval szigorúbb feltételeket szabnak a változók kapcsolatára vonatkozó társadalomelméleti feltevések statisztikai formalizálására vonatkozóan, és a modellezés célja robusztus állítások tétele egyes tényezők közötti oksági kapcsolatokra vonatkozóan. Ezzel szemben a gépi tanulási eljárások – melyek többségének célja a változók összessége alapján való előrejelzés – döntően adatvezérelt technikák, és éppen az elterjedt statisztikai modellekben jellemzően alkalmazott feltevéseket tesznek szükségtelemmé.

Valójában az utóbbi öt évben a társadalomtudományi statisztikában a két kutatási irány szorosan összekapcso-

lódott. Egyrészt az üzleti döntéshozatalt segítő, gépi tanulási eljárásokat alkalmazó adatelemzők körében jelentős az igény a döntések (kezelések) hatásának becslésére. Másrészt, a strukturális oksági modellezés mesterséges intelligencia adatvezérelt eljárásainak kritikáiból született; vezető kutatói jelenleg is az MI-kutatás legnagyobb hatású elméleti szakemberei között vannak. A kutatások egyik meghatározó iránya éppen a strukturális oksági modelleket is alkalmazó gépi gondolkodás (machine reasoning) fejlesztése. Továbbá az utóbbi 3–4 évben az oksági gépi tanulás legnagyobb hatású eredményeinek egy jelentős része ökonometriai kutatásokhoz és társadalomtudományi (ill. epidemiológiai) alkalmazásokhoz kötődik. Az új eljárások kidolgozását rengeteg elméleti és gyakorlati vita kíséri. Ezek egyrészt az adat- illetve elméletvezérelt mesterséges intelligencia mérnök kutatói, valamint az oksági gépi tanulás mérnöki és társadalomtudományi háttérrel rendelkező kutatói között zajlanak.

A kutatás célja kettős. Egyrészt a fent említett új kutatási irányok és eljárások, valamint a viták gyökereinek megértése és disszeminálása a TK és a MILAB kutatói között, és a hazai társadalomkutatói közösségben. Másrészt célunk, hogy a TK kutatói közösségének kvantitatív adatelemzéssel foglalkozó tagjai és a MILAB táblás adatok prediktív elemzésével foglalkozó kutatói megismerhessék és készségi szinten sajátíthassák el a fenti gépi tanulási technikák alapjait. A TK-ban 2020 óta létezik oksági adatelemzési olvasóköri. A kutatás első felében ezt folytattuk, a témához kapcsolható partnerek bevonásával és az eredmények disszeminálásával. Gépi tanulási továbbképzést és oksági statisztika előadásokat szerveztünk, továbbá statisztikai elemzésekkel vizsgáltuk azt a kérdést, hogy valóban megbízhatóbban becsülhetjük-e különböző kezelések hatását az új eljárásokkal, mint a hagyományos regressziós technikákkal.

Az oksági gépi tanulóval foglalkozó részprojektünkben egy adatgeneráló programot dolgoztunk ki. A projektünk következő munkaszakaszának célja egy társadalomtudományi és üzleti feladatokra optimalizált, később szolgáltatásként üzemeltethető *Datasheet* eszköz demo-verziójának kifejlesztése, amely oksági elemzések előtt ad segítséget kutatók és üzleti adatelemzők számára. A *Datasheet* a konkrét valós adatbázis tulajdonságai és a kutatói feltevések alapján generál szintetikus adatokat, és közli a különböző elemzési alternatívák tulajdonságait. A szoftver jövőbeni adatfelvételek tervezését is segítheti.

Eredményeink között megtalálható egy szabadon felhasználható, nyílt forráskódú rugalmas adatgeneráló algoritmus – önálló, módszer-tesztelő elemzésekhez és datasheet építéséhez; egy bevezető kurzusanyag (annotált kódok) gépi tanuláshoz társadalomtudósoknak; kölcsönös segítségnyújtáson alapuló együttműködés külső vállalati partnerrel innovációban valamint tanácsadás a Retail4, hazai közép vállalkozás által kifejlesztett kereskedelmi üzleti információs szoftver oksági adatelemzés moduljának létrehozásában.

### Résztevő kutatók

Buda Jakab  
Hajdu Gábor  
Janky Béla  
Szeitl Blanka  
Thamó Emese



# Mesterséges intelligencia képzési és ismeretterjesztési program

A Társadalomtudományi Kutatóközpontban a Mesterséges Intelligencia Nemzeti Laboratórium támogatásával átfogó képzési és ismeretterjesztési programot indítottunk az új elemzési technikák és a TK kutatási eredményeinek megismertetése céljából.

2020. november 3-án elméleti képzéssel, 2020. november 26-án pedig gyakorlati kurzus keretében megtartottuk az első kezdő szintű *Szövegbányászat és gépi tanulás R-ben* című oktatásunkat, melyeken különböző szervezetektől összesen közel 50 fő vett részt.
- Elkészült a Mesterséges intelligencia társadalomtudósoknak oktatási program alapjául szolgáló, saját korpuszaink elemzésein alapuló *Szövegbányászat és mesterséges intelligencia R-ben* (Sebők Miklós, Ring Orsolya, Máté Ákos) című tankönyvünk. Online konferenciával egybekötött könyvbemutatónk 2021. május 25-én került megrendezésre.
- Statisztikai elemzések gépi tanulási technikák segítségével – Python alapú bevezetés, címmel 4+4 alkalmas kurzust tartottunk, mely betekintést nyújtott a gépi tanulási eljárások logikájába és a legfontosabb eljárások algoritmusaiába (2021. október 15-december 10 között heti rendszerességgel, Python alapú mesterséges intelligencián nyugvó adatelemzési technikák oktatásával.)
2021. november 11-12-én ismételten alapozó kurzust tartottunk *Szövegbányászat és Mesterséges Intelligencia* képzési programunk keretében.
2022. február-április között nyolc alkalmas *Machine learning with Python* képzést tartottunk.
2022. május 26-án az alapozó szövegbányászati kurzusra épülve haladó szintű képzést tartottunk a *Szövegbányászat és Mesterséges Intelligencia* képzési programban.
2022. november 8-án és 15-én *Adatvizualizáció az R-ben* címmel tartottunk bevezető képzést és áttekintést az R ggplot2 csomagját használó adatvizualizációról.
2022. november 17-én és 24-én a Python programozási nyelv társadalomtudományokban való használatáról tartottunk alapozó kurzust.
2022. december 1-én második alkalommal tartottuk meg *Szövegbányászat és Mesterséges Intelligencia* – alapozó – képzési programunkat.

## További rendezvényeink

- Kutatásnépszerűsítő és képzési tevékenységünk keretében 2021. márciusától elindítottuk a *TK MILAB Speaker Series*, az *AI and Law International Research Seminar Series*, valamint a *Mesterséges intelligencia és a jog* online előadásorozatot (előadásonként 20-30 fő részvételével).
- 2020-21-es tanév tavaszi félévében több szakkollégiumi kurzust is tartottunk. (Bibó István Szakkollégium, Rajk László Szakkollégium).
- Mezei Kitti *Legal framework of autonomous vehicles* címmel 2021 őszén kurzust tartott a Budapesti Műszaki Egyetemen, valamint azonos címmel tananyagot is készített.
- Ambrus István és Mezei Kitti *Digitalizáció és büntetőjog* címmel kurzust tartott a Bibó István Szakkollégiumban szintén 2021 őszén.
2022. szeptember-október folyamán *Bevezetés az R használatába* kurzusunk során ismerkedhettek meg az R-ben való programozás és kvantitatív adatelemzés alapjaival a képzésen résztvevők.

## Impresszum

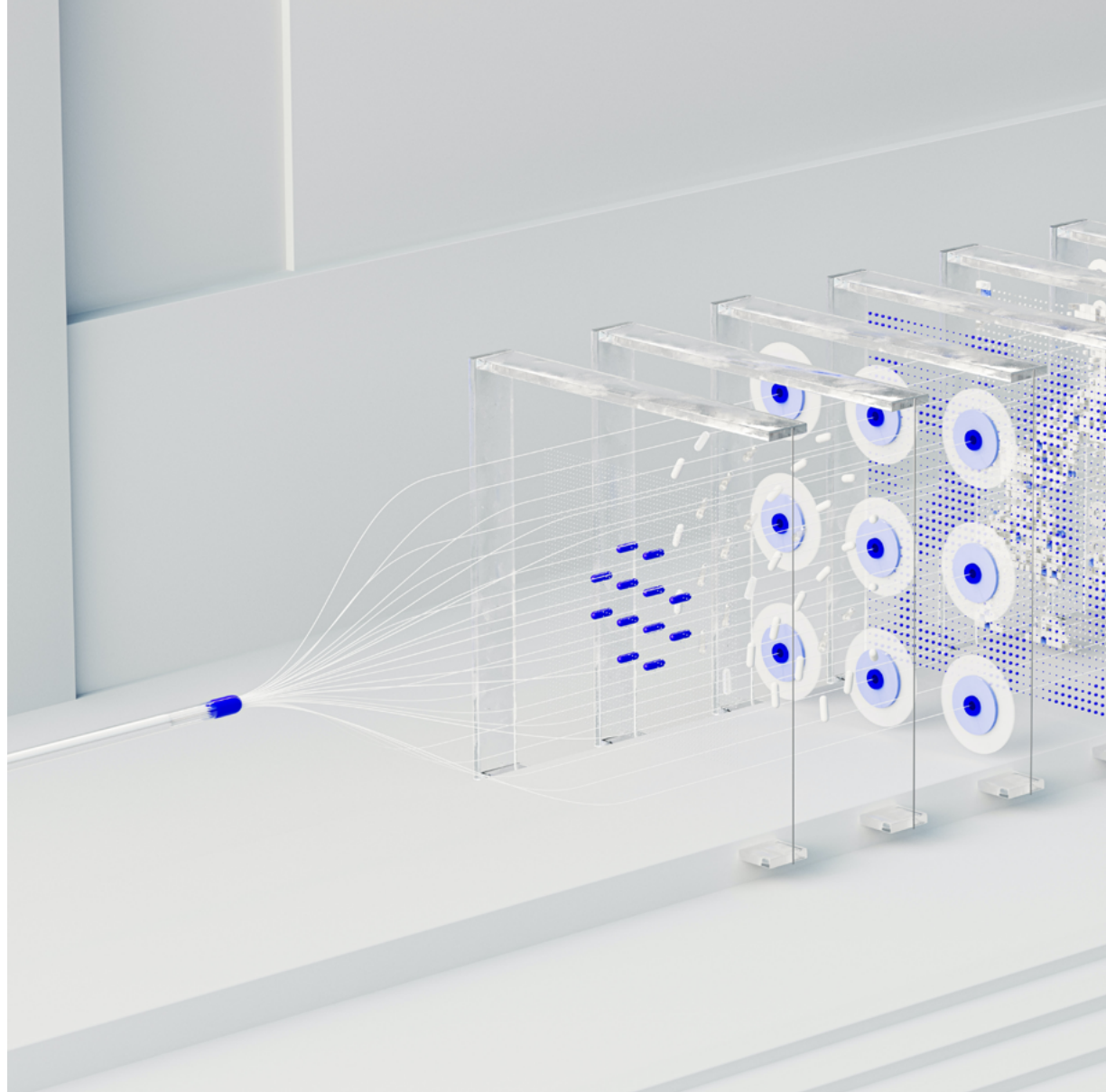
Készült az Európai Unió támogatásával,  
az RRF-2.3.1-21-2022-00004 azonosítójú,  
Mesterséges Intelligencia Nemzeti Laboratórium projekt keretében.

Kiadta a Társadalomtudományi Kutatóközpont 2023. februárjában.  
Felelős szerkesztő: Sebők Miklós, a TK MILAB projekt szakmai vezetője

Betűtípus: Moholy Sans és Segoe UI  
Kiadványszerkesztés: György Orsolya  
Fotók: Kovács Mariann, kivéve 1., 4. 17. és 22. oldal  
Borítófotók: Max Langelott és DeepMind, Unsplash  
Nyomta és fűzte a Másolófutár







**Széchenyi Terv  
Plusz** Építsük együtt  
Magyarországot!



MAGYARORSZÁG  
KORMÁNYA



Az Európai Unió  
társfinanszírozásával