Research accomplishments of the Artificial Intelligence National Laboratory at the Centre for Social Sciences (2020-2023)
I would like to welcome readers of this new publication of the Centre for Social Sciences. This volume presents research topics and results from the Artificial Intelligence National Laboratory. Due to a widespread use of AI and the impact on everyday life of the use of technologies, research into the social impacts of AI is becoming increasingly important. The participation of the Centre for Social Sciences in the work of the Artificial Intelligence National Laboratory is therefore of paramount importance. CSS researchers also rely on AI, as a technology, for their current social, legal, employment and policy-related research. The Artificial Intelligence National Laboratory can provide a professional framework for those efforts. I am convinced that this publication can offer exciting and interesting topics for everyone.
I am delighted to see this volume published, as it offers a glimpse into the wide-ranging efforts of the staff of the Centre for Social Sciences at the Artificial Intelligence National Laboratory. At first glance, it may not seem obvious to the outside observer what social scientists could possibly have to do with artificial intelligence. It must be remembered, however, that all issues of technology are ultimately social affairs: they function (or fail to function) in a particular social, economic and cultural context; their use triggers or implies social, i.e. attitudinal, behavioural, cultural or institutional changes; they raise ethical and legal dilemmas and the need for new regulations; they generate economic benefits for some actors while challenging other businesses; they create and destroy jobs; and so on. Add to this that methods based on artificial intelligence have also entered the toolbox of social researchers. In fact, I am proud to say that my colleagues not only use these methods, but have also contributed to their development. The present volume therefore provides an insight into the social aspects of the application of artificial intelligence, offering glimpses into the thematic richness of the relevant issues.

Zsolt Boda,
Director General, Centre for Social Sciences
Some decades ago, few would have imagined artificial intelligence and social science can have points of contact at so many areas and in such varied roles. Today, by contrast, it seems quite natural that engineers, scientists and social scientists use sophisticated AI-based analytical tools in their research, or work together to create effective, ethical and fair AI-based technologies that benefit society as a whole. As a result of the processes launched at the Centre for Social Sciences in recent years, artificial intelligence has come to the fore both in the methodologies employed (natural language processing, networks, big data, data donation, etc.) and in the research topics. A community has emerged of researchers able to view the theoretical traditions of social sciences from the perspective of this new technology. We sincerely hope that the AI research carried out at our institute can contribute to the development of science in Hungary, to the growth of social welfare and well-being and, last but not least, to further enhancing the presence of the CSS within the international academic community.
The social, economic and political impact of artificial intelligence in the 21st century marks a turning point in the history of humanity, without any exaggeration. There is a revolution comparable in importance to previous industrial revolutions: just as the spread of the steam engine or the use of electricity has led to the disappearance of old occupations and the emergence of new types of jobs, artificial intelligence is reshaping the labour market before our very eyes. The effects are positive and negative at the same time: the emergence of customised services and apps that serve billions of users is accompanied by breaches of privacy and even, in extreme cases, the abuse of personal data for political or commercial manipulation. As a founding member of the Artificial Intelligence National Laboratory, the Centre for Social Sciences aims to analyse the complex socio-economic and political impacts of the new technology, while developing new applications in language technology and other related fields. It has collaborated with dozens of national and international institutions, conducted interdisciplinary research projects, while also launching an educational program and a series of awareness-raising lectures involving hundreds of participants. The purpose of this publication is to provide an overview of the Centre’s far-reaching scientific activities in the field of artificial intelligence research.
An overview of CSS MILAB research projects

<table>
<thead>
<tr>
<th>Name of the project</th>
<th>Project leader</th>
<th>Work package</th>
</tr>
</thead>
<tbody>
<tr>
<td>8–9  AI-specific attitudes and narratives</td>
<td>Bence Ságvári</td>
<td>Sensor, IoT, Telecommunications</td>
</tr>
<tr>
<td>10–11 PoltextLAB – Text mining of political and legal texts using artificial intelligence</td>
<td>Miklós Sebők</td>
<td>Language technology</td>
</tr>
<tr>
<td>12–13 Octopus Research Tools (ORT)</td>
<td>Bence Ságvári</td>
<td>Sensor, IoT, Telecommunications</td>
</tr>
<tr>
<td>14–16 Legal challenges posed by artificial intelligence</td>
<td>Kitti Mezei</td>
<td>Security and personal data</td>
</tr>
<tr>
<td>17  Modelling the moral panic button (MPB) using traditional and NLP models</td>
<td>Endre Sik</td>
<td>Language technology</td>
</tr>
<tr>
<td>18  Measuring the communication of the central bank using large language models</td>
<td>Ákos Máté</td>
<td>Language technology</td>
</tr>
<tr>
<td>19  Hungarian benchmark databases for machine learning development</td>
<td>László Kiss and</td>
<td>Language technology</td>
</tr>
<tr>
<td></td>
<td>Csaba Molnár</td>
<td></td>
</tr>
<tr>
<td>20–21 What’s the shelf-life of a classification? A classification of anti-vax comments</td>
<td>Zoltán Kmetty</td>
<td>Language technology</td>
</tr>
<tr>
<td>22–23 Opinion analysis of political texts</td>
<td>Orsolya Ring</td>
<td>Language technology</td>
</tr>
<tr>
<td>24  A thematic exploration of textual research resources in CSS data repositories</td>
<td>Judit Gárdos</td>
<td>Language technology</td>
</tr>
<tr>
<td>25  Socio-cognitive effects in legal decision-making: a formal modelling</td>
<td>Gábor Péli</td>
<td>Language technology</td>
</tr>
<tr>
<td>26  Artificial intelligence and the use of drones in agriculture</td>
<td>Imre Kováč</td>
<td>Social and legal impacts</td>
</tr>
<tr>
<td>27  Work 5.0 – Technology and work in the age of AI</td>
<td>Miklós Illésey</td>
<td>Social and legal impacts</td>
</tr>
<tr>
<td>28  Causality machine learning: from principles to datasheets</td>
<td>Béla Janky</td>
<td>Social and legal impacts</td>
</tr>
</tbody>
</table>
The development of artificial intelligence (AI) will affect almost every aspect of our lives. Work, transport, health, education, justice and national security are all areas where technologies and processes are slowly but surely emerging with a result that more and more decisions or the preparation of decisions are carried out by AI-based systems.

However, this leaves us with a fundamental question: to what extent is this accepted by society? Altogether, the extent to which people are aware of the potentials offered by technology, their opinions on whether these solutions will make the world a better place or their concerns and mistrust about them are pivotal factors that will determine the developments in the near future. No successful AI-based solutions, whether in commercial or public services, can be created without a sufficient level of trust from users, i.e. society. That trust is based, most of all, on knowledge and information and the belief that these solutions will make our lives better, not worse.

That recognition has been embraced by Hungary’s National Artificial Intelligence Strategy, with one of its aims being to raise interest in AI among a large section of society, to familiarise people with the main features of AI, which should be accompanied by the communication and an awareness of the associated risks. Our research was primarily driven by the need for publicly available data on the domestic social environment related to AI in the context of that complex challenge.

In spring 2022, we therefore conducted a nationwide representative, face-to-face survey of 1,250 Hungarian respondents to find out the level of experience and awareness of the Hungarian adult population regarding the development and opportunities of digitalisation, in particular artificial intelligence, and their general expectations concerning AI-based technologies.

In the second step in the research series in January 2023, we were focusing on the ‘supply’ of information, as the information provided by the media
plays a crucial role in shaping opinions and knowledge about new technologies, including AI. So far, there have been few research efforts in Hungary to explore the way in which Hungarian online media platforms report on the uses of AI, its potentials and the latent and manifest messages online media conveys regarding the utopian vs. dystopian and trust vs. distrust dimensions. Our research analyses and interprets the news published and the relevant social media discourse between 2020 and 2022.

**Project participants**
Boxor Tamás
Pálvölgyi Eszter
Ságvári Bence

**Cooperating partners**
Corvinus University of Budapest
ZRI Závecz Research Kft.

**Research report**
I and AI – Values, attitudes and issues of trust concerning artificial intelligence in Hungarian society
The Text Mining of Political and Legal Texts (poltextLAB) project aims to employ Big Data methods for analyses of repositories of Hungarian and foreign political and legal documents. Traditional approaches to analysing qualitative data sources (texts, images and videos) have typically relied on a manual processing of data. While the knowledge of the source material remains essential in any social science study, there are obvious limitations to manual processing, particularly in terms of the reliability and validity of research results. The sheer variety and volume of data sources (e.g. a country’s entire body of legislation) can make manual processing impractical.

Quantitative text analysis and text mining approaches thus represent a new methodological standard for text-based Big Data projects in social sciences.

During the project, we build large text corpora, to be used primarily to develop and test the effectiveness of various machine learning algorithms based on artificial intelligence. In addition to developing new methodological solutions for the analysis of Hungarian texts, improving the effectiveness of existing algorithms and developing new hybrid methods for solving classification tasks, the project also aims to extend the methods to the analysis of non-Hungarian corpora. Various state-of-the-art major language models (e.g. BERT) have been successfully used in both Hungarian and multilingual classification, in up to 20 classes. At the end of 2022, we successfully launched our latest innovation, the CAP BABEL MACHINE, which relies on a multilingual BERT model for an automated identification of policy areas in texts, using the main topics of the policy code book of the Comparative Agendas Project (CAP). Through a form available at poltextlab.com/cap-machine, users can upload the files they wish to encode. After the encoding is finished, processed data is returned to users in a short time.

One of the priorities of the project is to set up a national and international network of like-minded researchers using text mining techniques. For years, we have been the main organiser of the international COMPTEXT conference, which aims to provide an opportunity for researchers relying on text mining techniques to hold regular meetings and learn about each other’s results. Our conferences are attended each year by around 150 participants from prestigious international universities and research institutions. As part of the project, 2 to 3 text mining training sessions are held for social scientists each year. Our Text Mining and Artificial Intelligence training program provides an introduction to supervised and unsupervised machine learning algorithms at both beginner and advanced levels. The program is built on our textbook Text Mining and Artificial Intelligence in R (authors: Sebők Miklós, Ring Orsolya, Máté Ákos), published in 2021, which is based on analyses of our corpora. Our Data Visualisation in R course aims to provide a practical and interactive overview of data visualisation using R’s ggplot2 package.

Cooperating partners
Jagiellonian University Kraków
National University of Ireland, Galway
University of Public Service
Graduate School for Public Administration
University of Pécs, Microsoft AI Knowledge Center
Reichman University
University of Szeged
University of Cologne
Publications


Project participants

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

Repositories

GitHub – poltextlab/textreuse_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

Multilingual comparable corpora of parliamentary debates ParlaMint 2.1

Major conferences

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

International grants awarded

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

Textbook of poltextLAB
Face-to-face scientific data collection has been facing a crisis all over the world in recent years. The methodology is becoming increasingly difficult and costly to implement. As a result, there is a growing interest in non-face-to-face data collection methods, including research focusing on users’ ‘digital footprints’, as indicators of human behaviour. Those data may derive from a number of sources, including website traffic, social media activity, smartphone app usage and data on movement and other physical conditions from various device sensors. Data is collected through non-intrusive methods, which do not require constant interaction from users, yet at the same time provide a high level of accuracy and detail about people’s digital lifestyles.

The Octopus project aims to combine the collection of sensor and device usage log-based digital behaviour data and questionnaires capable of displaying multimedia content within a complex, smartphone-based software ecosystem. This will result in an internationally innovative research tool for the scientific and even commercial investigation of a wide range of societal issues and problems, for modelling and predicting human behaviour and, on that basis, for supporting policy making. The development of the Octopus Research Tools started in autumn 2021 in the framework of the Artificial Intelligence National Laboratory, with the participation of social scientists and IT specialists. It was preceded by methodological groundwork and the specification of system functions. This involved an analysis and reconsideration of the features of software available on the market and developing new proprietary features.

A pilot study was conducted in 2021 on Hungary’s largest online research panel. A vignette technique was used to investigate patterns of respondents’ willingness to share app-based data, the role of research incentives, and the types of data people are more willing to or less willing to share or absolutely refuse to share for scientific, commercial and policy purposes. In the study analysing the research, we looked at the way people put a price tag on their own digital footprint compared to the personal information and opinions available through traditional questionnaires. Our results show that the content and context of data collection significantly affect people’s willingness to respond, while they associate different levels of risk and reliability with the different passive sensory data collection techniques. These methodological studies will continue in 2023, initially in the form of a pilot study with 100 respondents.
Heat map of smartphone usage intensity by minute over a 6-week period (aggregated user data)

Project participants
Gulyás Attila
Kollányi Bence
Koltai Júlia
Ságvári Bence

Publications

Cooperating partners
Nirprint Hungary Kft.
NRC Marketingkutató és Tanácsadó Kft.
SWT Informatika Kft.
TÁRKI Társadalomkutatási Intézet Zrt.
Legal challenges posed by artificial intelligence

Project Leader: Kitti Mezei

As part of the Artificial Intelligence National Laboratory, the project supports the implementation of the objectives of the National Artificial Intelligence Strategy. The project aims to identify current and future issues related to the regulatory environment of artificial intelligence (AI), map possible directions, identify AI-related good practices in different fields of law, relying on comparative methods and focusing on the EU legal framework. On that basis, specific and effective legal recommendations are put forward by researchers to promote a legal and regulatory environment for AI.

The specific characteristics of AI technology can make it difficult to monitor and effectively enforce compliance with existing international, EU or national legislation. According to the European Commission’s White Paper on Artificial Intelligence, the use of AI can lead to further breaches of fundamental rights, including the rights to freedom of expression, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, the protection of personal data and private life, consumer protection or the right to an effective judicial remedy and a fair trial. Those risks might result from flaws in the overall design of AI systems or from the use of data without correction. Security and design mechanisms should be integrated into these systems in order to ensure verifiable security. Research focuses on examining the legal and technical conditions for regulation by design.

The project aims to examine the Artificial Intelligence Act of the EU, which sets minimum horizontal requirements applicable to all AI systems placed on the market or used within the EU. The Artificial Intelligence Act follows a risk-based approach, whereby AI applications are classified into risk categories and the level of legal intervention required is adapted to the relevant risk level. In that context, the Act differentiates between uses of AI that create an unacceptable risk, high risk, medium risk and minimal risk.

Systems using AI are able to improve decision-making processes by analysing Big Data and exploiting its potential. A sufficiently large amount of such information, combined with the use of various data sources, may enable the identification of the data subject, resulting in a breach of data protection legislation. Like any other IT system, AI is exposed to cyber-attacks, which may result in a financial loss and even physical damage (e.g. in AI-based robots or AI-controlled critical infrastructure systems). Moreover, the increasing use of facial recognition technology is considered as the processing of ‘biometric data’ under the General Data Protection Regulation (GDPR) and is thus subject to strict conditions. The collection and use of such data for remote identification purposes, such as facial recognition in public places, pose risks to people's fundamental rights.

The research also focuses on the internal market-specific legal aspects of digital market regulation, in particular the role of AI. An increasing body of legislation has accumulated in the areas of competition, consumer protection and unfair trade in connection with the challenges posed by the increased use of AI in traditional and digital markets. The research examines the dynamics of EU regulation, specifically the regulation of AI and its impact on national legislation.

In addition to basic research, the project aims to assess sector-specific legal issues and provide professional guidance to consortium members. Apart from organising scientific and professional programs, a series of programs on Artificial Intelligence and Law has been launched as part of the project. In June 2022, in cooperation with the French Embassy and the French Institute in Hungary, the Institute for Legal Studies organised an international conference as a platform to discuss the current and future legal regulation of AI and its impact in Europe. The conference papers will be published in written form.
in 2023. The same year will also see the publication of a handbook on current regulatory issues in the field of AI, written by the researchers involved in the project, as well as a revised and expanded edition of Kitti Mezei’s monograph, first published in 2020. Various studies conducted by our researchers were also published in prestigious national and international journals during the project.

Project participants

AMBRUS István
CHRONOWSKI Nóra
CZEBE András
FRIEDERY Réka
GYÖRy Csaba
Hoffman István
HOFFMANN Tamás
KECSKÉS Gábor
KOVÁCS SZITKAY ÉsZTER
KÖRTvéLYESI Eszter
MEZEI Kitti
PAP András László
PAPP Mónika
POLYÁK Gábor
RÁCZ Lilla
SZALAY Klára
SZENTGÁLI-TÓTH Boldizsár
SZILÁGYI Emese
TRÁGER Anikó
VARJU Mátó
ZŐDI Zsolt

Publications


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 (Old confines, new challenges: the prudent use of AI in judicial work and its impact on the right to a fair trial). Glossa Iuridica, 8 (4), pp. 7-38. ISSN 2064-6887, 2022


KECSKÉS Gábor. Az autonóm járművek jogi kérdéseinek nemzetközi kontextusa, különös tekintettel a környezetjogi vetületekre (The international context of legal issues of autonomous vehicles, with particular reference to environmental aspects). Állam- és jogtudomány 61 (4), pp. 52-64., 2020

MEZEI Kitti. A kiberbűnözés aktuális kihívásai a büntető-jogban (Current challenges of cybercrime in criminal law). L’Harmattan Kiadó, Budapest, Magyarország, 282 p. 2021

MEZEI Kitti. A modern technológiák kihívásai a büntető-jogban, különös tekintettel a kiberbűnözésre (The challenges of modern technologies in criminal law, with a special focus on cybercrime). Állam- és jogtudomány 61 (4), pp. 65-81., 2020


MEZEI Kitti, 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; Józefów, pp. 73-80., 2022


RÁCZ Lilla. A személy és a dolog fogalmának (le-)hetséges változásai a mesterséges intelligencia és a kriptovaluták világában (Potential changes in the concepts of person and object in the world of artificial intelligence and cryptocurrencies). Állam- és jogtudomány 61 (4), pp. 82-107., 2020

SZENTGÁLI-TÓTH Boldizsár. A robotok jogi személyiségének koncepciója és annak lehetséges hatása a demokráciára: Miért ne adjunk szavazati jogot a mesterséges intelligenciáknak (Robotic personhood and its potential impact on democracy. Should artificial intelligence be citizens and vested with right to vote?). Magyar Jog 69 (6), pp. 321-330., 2022
Conference papers

ICON-S Mundo Conference – The Future of Public Law, own panel (AI and Law), online, 6–9 July 2021.


International conference organised

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

Cooperating partners

ELTE Faculty of Law
Curia
Mollia Zrt.
University of Public Service, Eötvös József Research Centre, Research Institute for Cybersecurity

International series of presentations

AI and Law program series

Online presentations by international visiting researchers:

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 (Brunel University, Law School)
Jakub Harasta (Masaryk University)
Rikke Jørgensen (The Danish Institute for Human Rights Copenhagen) Wen Xiang (University of Copenhagen)
Reijer Passchier (Open University/Leiden University)
The research aims to investigate the functioning of the Moral Panic Button (‘MPB’, a long-term manipulation technology created by the government) in online media and on Facebook. As part of the research, an existing database (a corpus extracted from SentiOne, consisting of articles and posts containing the keywords Soros and/or migration, written between February 2019 and February 2022) is analysed using both traditional methods and natural language processing (NLP) techniques. Within the database, various sub-corpora are also created of articles/posts related to the coronavirus epidemic within each subset. Various NLP techniques have been adopted to investigate the relationship between government-friendly and non-government-friendly actors and the narrative proxies generated by different combinations of the three keywords (Soros, migration, COVID-19, and their ‘synonyms’), as well as the likelihood of the existence of proxies (i.e. specific framing, specific terminology, character assassination, etc.) suggesting the existence of MPB.

Project participants
Buda Jakab
Farkas Eszter
Molnár Anna
Rakovics Szófia
Samoniek, Aleksandra
Stefkovics Ádám

Publications
Gerő Márton, Sik Endre. ‘Needn’t bother to press it now’ – the moral panic button (MPB) and the general elections of 2022. In Az állandóság változása, szerk.: Böcskei Balázs, Szabó Andrea, Gondolat/TK PTI, Budapest, pp. 65–86, 2022
Kerényi Péter, Sik Endre. Insight into the latest push of the moral panic button. Mozgó Világ, 11:15–27., 2022

Conference papers
Sik Endre. The Symphony of the Moral Panic Button – Prelude and Four Movements, HWK Fellow Lecture, 2021
While research on central bank communication is gaining prominence in economics and political economy literature, the typically employed dictionary-based techniques are far less efficient than supervised learning approaches. The project uses large volumes of ‘gold-standard’ quality teaching data to obtain a precise insight into the accuracy of certain dictionaries and the expected progress if supervised learning is used for classification. The supervised models include both long-established bag-of-words-based solutions (e.g. Naive Bayes, Random Forest, etc.) as well as state-of-the-art large language models based on deep neural networks (e.g. BERT, RoBERTa). The project will provide a comprehensive insight into how accurately a particular method is able to capture latent dimensions of central bank communication. This is particularly important as accurate measurement is a prerequisite for getting reliable results from the resulting econometric models. Preliminary results so far suggest that there are differences in magnitude between the models, which may also have a significant impact on the results of econometric models. The eventual fine-tuned language model will be open access and will be made available in the relevant data repository.

Project Leader: Ákos Máté

Project participants
BARCZIKAY Tamás
MÁTÉ Ákos

Publication

Conference paper

Repositories
GitHub – poltextlab/central_bank_communication: Replication materials
Script analysing the sentiment of central bank communication
R software package for the analysis of monetary sentiment

Economic sentiment in ECB presidential speeches
Hungarian benchmark databases for machine learning development

Project Leaders: László Kiss and Csaba Molnár

Due to the quick growth in digitalization and the rapid development of language technology tools, empirical and data-driven research directions aiming to address questions raised by social sciences by examining real text corpora have gained a prominent role in applied linguistics research in recent years. In this sub-project, we have improved CAP's existing databases and created benchmark databases by extending our existing data in two directions. We have included (oral) questions alongside the actual questions to ministers, quick-fire questions and speeches before the order of the day between 2010 and 2022 on the one hand, while adding the texts from the two parliamentary cycles between 2002 and 2010 to our existing data and corpora for those four types of texts.

Substantial manually or machine-annotated corpora may also be a suitable resource for state-of-the-art techniques such as the training of machine learning models. While almost all of the most advanced such models are currently based on the use of artificial neural networks, due to reasons of efficiency (i.e. the use of large models such as BERT is often a very compute-intensive process, not only in terms of training but also in terms of its application during prediction), the use of earlier methods (logistic regression, support vector machine, etc.) are still justified.

During the research, machine learning experiments were carried out on the corpora created in order to assess their practical value in addressing sociological issues. A wide spectrum of methods (e.g. SVM, LSTM, BERT, etc.) have been employed in order to assign the most appropriate technology to each research question. In addition to the benchmark database built during the research, the Parlawspeech database that includes the corpus and data of parliamentary speeches, bills and laws between 1994 and 2022, the database of Magyar Nemzet front pages between 2002 and 2014, have been created, and a study entitled If there is nothing else to say: the local content of interpellations by Csaba Molnár, has published in the Journal of Legislative Studies.

Project participants

Kiss László
MOLNÁR Csaba
BARCZIKAY Tamás
Kiss Rebeka
KLEIN Adrienn
KOVÁCS Viktor
KUBIK Bálint György
Pokorny Zsanett
ÜVEGES István

Publications

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

Conference paper


Repository

Hungarian PARLAWSPEECH dataset
As a larger goal, our project focusing on COVID-19 anti-vax sentiment has aimed to create a platform to support policy-making that is able to draw on textual data generated in the online space to provide a real-time picture of the opinions and reactions of the Hungarian population. Most of the classification projects have been classifying data sets that have been completed in time. However, an ever-increasing body of data is made available by the dynamic growth in digitalization, which may even require an ongoing classification process to be put in place. Such continuity is another important feature of our project. While certain classification models are able to continuously retrain themselves on the basis of new input data, such adaptation is more difficult in the case of models based on training datasets. As the cost of retraining models is high, it is important to consider how long you intend to use a model before replacing it.

In our sub-project, we developed a framework between 2020–2021 to monitor the discourse on vaccination and vaccine scepticism in the online space in Hungary. One of the core elements of this framework is the BERT model which classifies comments according to the degree of anti-vax sentiment expressed in them. The initial model was completed in spring 2021, based on the manual classification of 10,000 comments. As the data was collected on an ongoing basis, during which time the discourse continued to evolve, this raised the issue of the temporal stability of the model. To examine this question, we prepared a new annotation at the end of 2021, based on the comments from autumn 2021.

The new annotation enables the examination of the performance of different annotations at different times, whether it is worth treating annotated data as a single unit, and the efficiency of special techniques such as rewriting comments into a different category. During 2022, we primarily focused on the extent to which the classification of comments can be improved by the use of unlabelled data, from which periods it may be worth including unlabelled data, and whether it makes sense to pre-select content on the basis of certain criteria during the annotation process.
The platform developed under the MILAB project tracked in real-time how and when each vaccine related to COVID vaccination appeared in the online space.

**Weekly percent change in number of posts**

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Weekly percent change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfizer</td>
<td>29.7%</td>
</tr>
<tr>
<td>Sinopharm</td>
<td></td>
</tr>
<tr>
<td>Sputnik</td>
<td></td>
</tr>
<tr>
<td>Astraeneca</td>
<td></td>
</tr>
<tr>
<td>Moderna</td>
<td></td>
</tr>
</tbody>
</table>

**Weekly percent change in vaccine mentions in number of posts**

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Weekly percent change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfizer</td>
<td>91.9%</td>
</tr>
<tr>
<td>Sinopharm</td>
<td>193.8%</td>
</tr>
<tr>
<td>Sputnik</td>
<td>278.0%</td>
</tr>
<tr>
<td>Astraeneca</td>
<td>318.2%</td>
</tr>
<tr>
<td>Moderna</td>
<td>-100.0%</td>
</tr>
</tbody>
</table>

**Weekly percent change in vaccine mentions in number of comments**

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Weekly percent change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfizer</td>
<td>42.6%</td>
</tr>
<tr>
<td>Sinopharm</td>
<td>23.5%</td>
</tr>
<tr>
<td>Sputnik</td>
<td>65.8%</td>
</tr>
<tr>
<td>Astraeneca</td>
<td>150.0%</td>
</tr>
<tr>
<td>Moderna</td>
<td>68.8%</td>
</tr>
</tbody>
</table>

---

**Project participants**

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

**Publication**


**Cooperating partners**

SentiOne
Ynsight

**Conference papers**

Kmetty Zoltán. Investigating the evolution of COVID-19 vaccination-related online discourse using text mining techniques. 8th Education and Research Methodology Workshop, Semmelweis University, 3 February 2021.


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, 27–31 July 2021.

**Demo**

MILAB COVID Dashboard

---

The platform developed under the MILAB project tracked in real-time how and when each vaccine related to COVID vaccination appeared in the online space.
Opinion analysis of political texts

Project Leader: Orsolya Ring

The main objective of the project is to develop sentiment and emotion analysis procedures for the analysis of various types of Hungarian texts (online news portals, newspaper articles, political speeches and parliamentary speeches). From each text analysed, the procedures employed aim to extract the pieces of information expressing an assessment. The analysis can be conducted at different levels, depending in part on the basic unit of analysis and whether the object or the trigger of emotion are determined. The literature distinguishes between sentiment analysis along a positive-negative-neutral scale and emotion analysis based on multiple categories. The latter provides much more information on the emotional charge of a given unit.

We have been experimenting with a number of methods to achieve our goal. We have created substantial manually annotated sentiment and emotion corpora, which are then used for dictionary construction, machine training, and testing the effectiveness of machine learning and dictionary-based parsing algorithms. As part of the project, we have created a sentiment dictionary that works along a positive-negative scale, applying a word embedding technique to text published on online news portals. We have developed an inductive emotion categorisation system that is able to distinguish between 12 different emotions in political texts. As our system of categories is convertible to internationally used emotion categories, it is suitable for comparison with such categories.

Using a double-blind coding technique, we constructed a sentiment and emotion corpus of 5,700 sentences annotated at sentence level, and a sentiment and emotion corpus of parliamentary speeches annotated at sub-clause level (HunEmPolli), where the 39,840 identified emotions were linked to the relevant arguments. Both our corpora were annotated under strict quality control and with great unanimity among coders. In the current project phase, we are using training data generated from the HunEmPolli corpus and are fine-tuning the huBERT model to train a sentiment and emotion analysis model. We have also created a website to visualise our parliamentary speech corpus: https://napirendek.hu/erzelmek/.

Project participants
Guba Csenge
Ring Orsolya
Szabó Martina Katalin
Váradi Bendegúz
Vincze Veronika

Publications
Ring Orsolya, Vincze Veronika, Guba Csenge, Üveges István. HunEmPolli: magyar nyelvű, részletesen annotált emóciókorpusz (HunEmPolli: Hungarian emotions corpus with detailed annotations). In: Berend Gábor, Gosztolya Gábor, Vincze Veronika (szerk.) 19th Conference of Hungarian Computational Linguistics, University of Szeged, Information Technology Institute, 2023
Szabó Martina Katalin, Vincze Veronika, Ring Orsolya, Guba Csenge. Nagyot mondó képviselők? Fokozás a politikai kommunikációbán (MPs telling tall tales? Stepping up emphasis in political communication). In: Berend, Gábor; Gosztolya, Gábor; Vincze, Veronika (eds.) 18th Conference of Hungarian Computational Linguistics, University of Szeged, Information Technology Institute, 2022
Ring Orsolya (Erjavec, Tomaž, Ogrodniczuk, Maciej, Osenova, Petya et al.) The ParlaMint corpora of parliamentary proceedings. Language Resources and Evaluation, 2022
Conference papers
Szabó Martina Katalin, VINCZE Veronika, Ring Orsolya, Guba Csenge. MPs telling tall tales? Stepping up emphasis in political communication. 18th Conference of Hungarian Computational Linguistics, online, 27–28 January 2022.

Cooperating partners
Budapest University of Technology and Economics, Department of Telecommunications and Media Informatics, SmartLAB
Charles University, Prague
Kempelen Institute, Bratislava
Montana Tudásmenedzsiment Kft.
Vistula University, Warsaw

Repository
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
HunEmPolici corpus – Github
Possibilities and limitations of a lexicon-based sentiment analysis of Hungarian political news – Github
A thematic exploration of textual research resources in CSS data repositories

Project Leader: Judit Gárdos

Our pilot project aimed to improve archive searchability by testing various machine text analysis techniques on a sample of interview collections held in the digital social science archives of the Voices of the 20th Century and the Research Documentation Centre (KDK) of CSS. The project was carried out in cooperation between the KDK and SZTAKI’s Department of Distributed Systems. After selecting, applying and validating the most appropriate technique, results were integrated into the beta version of a repository search engine. In effect, metadata were automatically assigned to each interview, which provide the researchers with information about the content of the texts (interviews or extracts of interviews) and the location of the texts and text excerpts that are related to each other (even across several collections) and are relevant to the research issues.

Subject headings and subject indexes were generated for the interview texts, manually first and then aided by machines. Their adequacy was verified by our researchers. Even in those cases where documents were subject-ed to machine analysis only, without subsequent manual verification, the validation of results has led to efficient subject heading and subject index generation. Rather than being simply keywords or synonyms of keywords appearing in the texts, the subject headings or labels associated with the interviews are elements of a conceptual network created using the ELSST, an international social science thesaurus, which helps to reveal the sociological phenomena inherent in the texts. Steps have also been taken in the direction of NER (Named Entity Recognition). We identified and then wikified name elements and time tags appearing in the texts, linking them to the Wikidata knowledge graph, Geonames, VIAF, PIM and other name spaces.

To improve searchability, abstract subject headings and name elements obtained through machine processing are associated with the documents in several new metadata fields. Existing documents are thus opened up for new research. The subject headings are translated into English, making our archives searchable for researchers abroad. As a result, domestic resources that have previously been inaccessible due to the language barrier are now becoming visible and accessible to the international research community. The results of machine processing are visualized by highlighting the name elements and linking them to the dictionary entries as well as by exploring and displaying the frequency of certain topics and subject headings and their relationship.

On this project, we worked in cooperation with CESSDA (Consortium of European Social Science Data Archives). As part of that cooperation, a Hungarian translation of the ELSST English-language social science thesaurus, containing more than 3,300 terms, was completed in collaboration with the Research Centre of Linguistics (NYTK). It has been available online since September 2022. The project also involved a cooperation with the Budapest University of Technology and Economics (BME) for improving the efficiency of BEAST, a Hungarian database speech transcriber. BEAST is an open-source, research-ready system based on the SpeechBrain code, developed by the NYTK and BME, with financing from the Hungarian Scientific Research Fund (OTKA) and MILAB, which uses state-of-the-art transformer neural structures.

Researchers interested in the sociological sources will be able to access the results of our exploration of interview documents on a common online search platform created for the repositories of the CSS Research Documentation Centre.

Project participants

Publication
Egyed-Gergely Júlia, Vajda Róza, Gárdos Judit, Horváth Anna, Meiszterics Enikő, Micsik András, Martin Dániel, Marx Attila, Pataki Balázs, Siket Melinda.

Szociológia, kutatási adatok, mesterséges intelligencia: lehetőségek és tapasztalatok (Sociology, research data and artificial intelligence: opportunities and experiences).


Conference papers
FAIRsFAIR (EOSC alprojekt) Final Event. The National Perspective, online roundtable talk, Gárdos Judit, 26 January 2022.

Socio-cognitive effects in legal decision-making: a formal modelling

Project Leader: Gábor Péli

The evolution of the fuzziness of legal concepts, explored by our research, is an area that has so far resisted any attempts at formal modelling. Making conceptual ambiguities measurable and understanding their evolutionary dynamics help expand the toolbox for achieving coherent and predictable decision-making in the functioning of the legal system. The semantic variants of legal concepts are modelled by density functions over the semantic space, whereas the dynamics of legal ambiguity is modelled by variations in the shapes and spatial distances of such functions. Similar formal methodological approaches are being developed in several affiliated research centres abroad, for example by analysing the ambiguity in conceptual classification patterns in various fields of art. These foreign research efforts are also in the testing phase of such field-specific theoretical frameworks. Therefore, our research aims to set up the domestic leg of an emerging new consortium, to which we would contribute by analysing the dynamics of legal concepts. Formal modelling and the manual collection of empirical material are to be followed by contextual Deep Learning tests.

In the initial phase of the research, we investigated the semantic variants of basic legal concepts and identified their characteristic cognitive dimensions. To that end, we analysed detailed assessments of legal concepts by the Constitutional Court, looking for specific descriptive dimensions. The experiences drawn during the first phase on mapping into cognitive space were applied to second-instance and Curia-level rulings on domestic criminal judgments (pilot study). The aim was to operationalize concepts and clarify their cognitive spatial dimensions. A presentation on the results of the exploratory phase of the research and the consequent adaptation of our initial formal model was held at an international conference in Berlin in June 2022.

In addition to the above, we manually collected data on the frequency of occurrences of the semantic variants of concepts in the relevant cognitive framework in the criminal law database. Understanding the probabilities of occurrence allows estimates of the cognitive distances of legal concepts (a relative entropy calculation based on the Kullback-Leibler divergence). The empirical study will be published by submitting two articles in English and an article in Hungarian to a law journal. Further research materials are produced while carrying on with manual coding. As we begin to prepare for machine learning-based coding, we draw on parallel research by participants at the Berlin conference, which have adapted the same formal cognitive model to contexts not related to legal concepts.

Project participants
Közantal Viktória
Lőrinicz Viktor
Péli Gábor
Pólos László

Conference paper
Artificial intelligence and the use of drones in agriculture

Project Leader: Imre Kovách

Precision farming is an essential element of agricultural innovation. Its most important feature is the emphasis on the role of the geographical information system (GIS) and remote sensing solutions, accurate measurement and computer control in all aspects of farming. Precision agriculture is the most important example of the emergence of information technology and artificial intelligence in agriculture. It represents a revolution in agriculture, driven by an effort to improve the profitability and efficiency of farming in harmony with the environment, in order to achieve social benefits. Precision agriculture also pursues social objectives: it helps to improve food quality and safety, while contributing to a better management of climate change and natural resources by reducing environmental pressures. The use of AI-based technology is environment- and soil-friendly, as well as a great tool to improve profitability.

The two research topics are the social conditions for the adaptation of precision farming (e.g. precision mapping of soil quality and yields) and a network analysis of its growth. It is a basic challenge for research on social conditions to identify the indicators of the characteristics of willingness by each group of farmers, i.e. innovators (inventors), early adopters, early majority, late majority and laggards, to adopt precision technologies, and the factors potentially influencing their motivations to switch to AI-based precision farming. The second research topic is an analysis of the network components of the growth in precision farming. This includes research on the diffusion of information on the method and the degree to which this is determined by the network, and a network analysis of the increase in partial or full conversion to precision farming.

Between December 2021 and the end of January 2022, we conducted a questionnaire-based survey of 200 precision farmers. The sampled farms are representative of Hungarian crop farms both regionally and in terms of scale. The survey extracted detailed information on the use of different forms of AI. In particular, the questions focused on the motivations and practical aspects of the use of drones, which is one of the most recent AI tools in agriculture, spreading more rapidly than other kits. The questionnaire survey is complemented by 30 audio interviews of precision farmers. The first analysis of the research focuses on farmers’ motivations for using drones. Due to the adoption of a trans-theoretical model (ordinal logit regression model) and the structure and questions of our questionnaire, our results are suitable for comparison with a German study conducted in the field. Many of the German questions were included in our questionnaire. Information from qualitative interviews is also used to interpret the results of our mathematical and statistical analysis. The research results are suitable for direct practical application, revealing both the nature and the combined impact of factors encouraging the use of drones.

Project participants
Bai Attila
Balogh Péter
Czibere Ibolya
Gabnai Zoltán
Kovách Imre
Loncsák Noémi
Megyesi Boldizsár
Nemes-Zambo Gabriella

Publications
Bai Attila, Kovách 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), 2020


Cooperating partners
University of Debrecen
Kynetec Hungary Kft
In recent years, research into digitalisation and artificial intelligence and their societal impact has intensified, and so has the debate on the relevant research results. The significant productivity gains expected from the adoption of advanced smart technologies are as of yet unrealised. One of the most important possible explanations is that the expected joint optimisation of technological and social systems has failed to take place as the evolution of the latter has lagged behind technological progress. The idea is nothing new: it was studied at London’s Tavistock Institute as early as during World War II. According to the theory emerging from that research, it is of paramount importance in corporate practice to align technology, people and the work organisation. This idea is called the theory of sociotechnical systems design.

As part of our research, we conduct secondary analyses on the databases of large-scale European questionnaire surveys (European Company Survey, European Working Conditions Survey). Answers are sought to the following questions: what is the position of Hungarian businesses in the practice of digital transformation, how widespread so-called learning organisations are in Hungary, to what extent Hungarian businesses rely on the knowledge of their employees and what tools they use to mobilise that knowledge, and what kind of correlation exists between digital transformation, employee participation and corporate innovation performance.

By answering those questions, the analyses seek to explore the impact of digital technologies on the work organisation and the importance of socio-organisational (‘soft’) skills in work practice (management practice), such as communication and cooperation skills, the development of individual and collective knowledge, the ability to solve complex tasks, or negotiation and persuasion skills. By analysing European databases, we will be able to investigate these issues from a perspective of international comparison, allowing a comparison not only with the European average, but also with the Visegrad group of countries, which have followed a similar development path.

International online workshop organised

Knowledge Infrastructure in the Platform Economy, online, 21 April 2022. Organised in cooperation with the University of Public Service, with participants from the US, Portugal, Serbia, Germany, Finland, Australia and India, on the uses at the workplace and impacts of artificial intelligence and algorithmic management.

Project participants

ILLÉSSY Miklós
Makó Csaba
Saeed Nostratabadi
Pap József

Publications

ILLÉSSY Miklós, HUSZÁR Ákos. Technológiai fejlődés és munkaerőpiac: automatizációs kitettől Magyarrországon (Technological development and the labour market: how susceptible are jobs to automation in Hungary?) 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


Conference papers


Causality machine learning: from principles to datasheets

Project leader: Béla Janky

Over the last decade, two areas of social statistics have seen research that has had a significant impact, bringing about a paradigm shift in statistics. Structural causal modelling, on the one hand, is breaking new ground in social research practice, while the application of machine learning techniques in engineering is revolutionising econometrics. The two areas mark seemingly opposite paths of development for social data analysts. The criteria imposed by structural causality models on the statistical formalisation of social theoretical assumptions concerning relationships between variables are considerably stricter than current practice, while modelling aims to put forward robust claims about the causal relationships between various factors. By contrast, machine learning techniques, which mostly aim to offer predictions based on a set of variables, are primarily data-driven techniques, rendering unnecessary the very assumptions that are typically used in widely applied statistical models.

Over the last five years, the two research directions have in fact become closely intertwined in social statistics. First, there is strong demand among data analysts relying on machine learning techniques to support business decision-making to estimate the impact of decisions (treatments). Second, structural causal modelling emerged out of the criticism of data-driven methods in artificial intelligence in the first place; its leading researchers are currently among the most influential theorists in AI research. One of the key directions of research is the development of machine reasoning, which draws on structural causal models. Moreover, a significant part of the most influential results of causal machine learning in the last 3 to 4 years have been related to econometric research and social science (or epidemiological) applications. The development of new procedures is accompanied by considerable theoretical and practical debate. These tend to take place between engineer researchers in data-driven or theory-driven artificial intelligence, and researchers in causal machine learning with a background in engineering and social sciences.

The purpose of the research is twofold. We aim to understand the new research directions and processes referred to above and the root causes of debates and to disseminate our conclusions among researchers of CSS and MILAB as well as within the national social research community. Another goal is to enable members of the CSS research community specialising in causal machine learning with a background in engineering and social sciences. The software can also aid the design of future data collection.

We also developed a data generation program as part of our sub-project on causal machine learning. The next phase of our project focuses on the development of a demo version of a Datasheet tool optimised for both social science and business tasks, which can later be operated as a service to assist researchers and business data analysts prior to carrying out causal analyses. Generating synthetic data based on the properties of a specific existing database and researchers’ assumptions, the Datasheet communicates the properties of various analysis alternatives. The software can also aid the design of future data collection.

Our results to date have included a free-to-use, open-source flexible data generation algorithm for stand-alone method-testing analyses and datasheet building; an introductory courseware (annotated codes) on machine learning for social scientists; mutual collaboration on innovations with an external corporate partner; and consulting on the development of a causal data analysis module for the Retail4 commercial business information software developed by a medium-sized Hungarian company.

Project participants
Buda Jakab, Hajdu Gábor, Janky Béla, Szeitl Blanka, Thamó Emese
Artificial Intelligence training and education program

A comprehensive training and popular education program has been launched by the Centre for Social Sciences, with the support of the Artificial Intelligence National Laboratory, to promote new analytical techniques and to disseminate the Centre’s research results.

The theoretical and practical sessions of our first basic-level training on *Text Mining and Machine Learning in R* were held respectively on 3 and 26 November 2020. They were attended by nearly 50 participants from various organisations.

Various special courses were organised in the spring semester of the 2020/2021 academic year (Bibó István College for Advanced Studies, Rajk László College for Advanced Studies).

As part of our research and training activities, in March 2021 we launched the CSS MILAB Speaker Series, the AI and Law International Research Seminar Series and the online AI and Law lecture series (for an audience of 20 to 30 participants at each lecture).

A textbook entitled *Text Mining and Artificial Intelligence in R* (Miklós Sebők, Orsolya Ring, Ákos Máté), based on analyses of our corpora, is now available and used as a basis for the Artificial Intelligence for Social Scientists training program. Our book launch event, organised alongside an online conference, took place on 25 May 2021.

We held a 4+4-session course entitled Statistical analyses using machine learning techniques – A Python-based introduction to provide insight into the logic of machine learning processes and the algorithms of the most important techniques (weekly sessions between 15 October and 10 December 2021 on Python-based artificial intelligence data analysis techniques).

In autumn 2021, a course entitled Legal framework of autonomous vehicles was held by Kitti Mezei at the Budapest University of Technology, while she also produced a course material under the same title.

Also in autumn 2021, István Ambrus and Kitti Mezei held a course at the Bibó István College for Advanced Studies on Digitalisation and Criminal Law.

On 11 and 12 November 2021, we held another foundation course as part of our Text Mining and Artificial Intelligence training program.

Eight sessions of a Machine learning with Python course were held between February and April 2022.

Building on the text mining foundation course, we held an advanced training course under the Text Mining and Artificial Intelligence program on 26 May 2022.

In September and October 2022, we held an Introduction to R course to introduce participants to the basics of programming and quantification in R.

The introductory workshop Data Visualization in R held on 8 and 15 November 2022 provided an overview of data visualization using the R ggplot2 package.

On 17 and 24 November 2022, we held a basic workshop on the uses of the Python programming language in social sciences.

The second edition of our Text Mining and Artificial Intelligence foundation training program was held on 1 December 2022.
Publisher's imprint
Produced with support from the European Union, as part of project RRF-2.3.1-21-2022-00004 of the Artificial Intelligence National Laboratory

Published by the Centre for Social Sciences, February 2023
Managing editor: Miklós Sőbö, Co-Principal Investigator of the CSS MILAB project

Fonts: Moholy Sans és Segoe UI
Design and layout: Orsolya György
Photos: Mariann Kovács (CSS), except in pages 1, 4, 17 and 22
Cover photos: Max Langelott and DeepMind, Unsplash
Printed by Másolófutár