



BALTIC JOURNAL OF LAW & POLITICS

A Journal of Vytautas Magnus University

VOLUME 15, NUMBER 1 (2022)

ISSN 2029-0454



Cit.: *Baltic Journal of Law & Politics* 15:1 (2022): 141-172

<https://content.sciencedirect.com/view/journals/bjlp/bjlp-overview.xml>

DOI: 10.2478/bjlp-2022-0007

DIGITAL TRANSFORMATION OF LEGAL SERVICES AND ACCESS TO JUSTICE: CHALLENGES AND POSSIBILITIES

Julija Kiršienė

Professor, Ph.D.

Vytautas Magnus University, Faculty of Law (Lithuania)

Contact information

Address: Jonavos g. 66, LT-44191, Kaunas, Lithuania

Phone: +370 37 751044

E-mail address: julija.kirsiene@vdu.lt

Darius Amilevičius

Associate Professor, Ph.D.

Vytautas Magnus University, Faculty of Informatics (Lithuania)

Contact information

Address: Vileikos g. 8, LT-44404, Kaunas, Lithuania

Phone: + 370 37 327900

E-mail address: darius.amilevicius@vdu.lt

Dovilė Stankevičiūtė

Ph.D. Student

Vytautas Magnus University, Faculty of Law (Lithuania)

Contact information

Address: Jonavos g. 66, LT-44191, Kaunas, Lithuania

Phone: +370 37 751044

E-mail address: dovile.stankeviciute@vdu.lt

Received: May 16, 2022; reviews: 2; accepted: July 26, 2022.

ABSTRACT

The pandemic affected the access to justice situation in terms of the never rapid shift to digitalisation of legal services, and in this article, we evaluate whether artificial intelligence (AI) and its state-of-the-art technologies like machine learning and human language technologies have the potential to improve access to legal services. For this purpose, we not only examine and identify problematic areas, but also share the empirical data and insights of the practical application of AI technologies, especially human language technologies. In the first part of the article, we explore how the internet has created the foundations for a new paradigm of society including institution law. The second part of the article is devoted for analysis of challenges for access to justice in post pandemic world. In the third part, we elaborate on questions about technical feasibility, legal and moral acceptability of the digitalisation of legal services. Then follows the case analysis of the practical application of human language technologies in legal domain.

KEYWORDS

Digitalisation, access to justice, human rights, legal services, artificial intelligence, human language technologies, legal language

NOTE

This research was funded by the European Social Fund under the Improvement of Researchers Program for implementing world-class R&D projects according to Measure No. 09.3.3-LMT-K-712.

INTRODUCTION

The COVID-19 pandemic has exacerbated the problem of limited access to justice even though during this crisis the need for legal services has grown. Despite progress in access to justice policy implementation measures in many countries, millions of people even in rich economies still lack sufficient and adequate access to legal services.

In this article, we explore how the pandemic affected the access to justice situation in terms of the never rapid shift to digitalisation of legal services, and we evaluate to what extent technological advancements, such as artificial intelligence¹ (AI) and its state-of-the-art technologies like machine and deep learning and human language technologies² (HLT), can impact the overall productivity of justice systems and are capable of automating the legal reasoning process in such a way as to improve access to justice and legal services. Such breakthrough technologies as AI and its subfields, such as machine and deep learning, HLT, and data analytics, are already quite widely used. Process automatization in areas like law, health care, education, and decision-making in the public sector is not possible without the full use of state-of-the-art digital language technologies. However, the use of the above-mentioned technologies in legal services in many countries (for example, Lithuania) is quite low, and therefore changes in this area are absolutely necessary and could have a significant impact on achieving better access to justice. So, paradoxically, although breakthrough technologies are applied in many areas and show surprising results, the application of these technologies in the field of law (not only in Lithuania, but also all over the world) remains sluggish. Preliminary research shows that this is mainly affected by the following factors: human (fear of losing traditional work conditions, lack of access to technology, insufficient technological literacy, etc.), legal (unclear and/or insufficient legal regulation), and technological (complexities of language [e.g., Lithuanian] digitalization, legal language specifics, a lack of datasets for research and training machine and deep learning models, etc.).

The relevance and novelty of this research are justified by the unprecedented challenges of recent years for the world, in particular the COVID-19 global pandemic. These challenges also caught the legal services sector unprepared. Moreover, the use

¹ Emerging era of AI, the Internet of Things applications, and the mutually beneficial relationship between the Internet of Things and Artificial Intelligence are enabling disruptive innovations and bringing AI closer to smart technologies. Due to the limits of the paper, we will use the terms "AI technologies" and "smart technologies" interchangeably. For further consideration on "smart", see Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham, UK: Edward Elgar, 2015).

² "Human Language Technologies" refers to the branch of AI that enables computers to analyse, interpret, and manipulate human language. Because their subfields continue to grow and contain the abbreviation "Natural Language", they are also known as "Natural Language Technologies": "Natural Language Processing", "Natural Language Understanding", "Natural Language Generation", "Natural Language Query", "Speech technologies".

of breakthrough technologies in the search for opportunities to increase access to justice has not been studied thoroughly so far, taking as examples specific countries and their legal systems and language peculiarities. Due to the dynamics of technology, even research that has been carried out is losing its relevance very quickly. For this reason, an analysis of the development of a standard for access to legal services with a strong focus on digital technological advancements follows, drawing on the best practices from the highly dynamic areas of AI technologies.

For this purpose, in this research, we not only examine and identify problematic areas, but also share the empirical data and insights of the practical application of HLT to improve the understanding of challenges that may occur when applying AI in the legal domain. This research may also lead others to reconsider the practices of access to other public services like health, education, finance, and others. As an empirical example, we most often refer to the situation in Lithuania and the Lithuanian language as a technologically challenging example for AI in the legal domain; however, the insights from this research can be applied to other languages, especially those with fewer resources in terms of digitalization than the English language.

1. NEW PARADIGM OF LAW IN TRANSFORMING DIGITALIZED SOCIETY

In terms of technological advancement, the post-pandemic world is shifting from the age of digital electronics, which some describe as the third industrial revolution,³ to the fourth industrial revolution, which is characterised by a fusion of technologies.⁴ Currently, a further shift—from Industry 4.0 to Industry 5.0—is being widely discussed. It will be distinguished by the collaboration between machines and human beings,⁵ merging physical and cyber spaces. Although some argue that “society and technology develop in tandem,”⁶ it is also obvious that such technologies as the internet, AI, robots, blockchains, etc., which are blurring the lines between human

³ A short discourse on the history of industrial revolutions: the First industrial revolution used water and steam power to mechanise production; the Second used electricity for mass production; the Third used electronics and information technology to automate production; the Fourth is characterised by a fusion of technologies that blurs the boundaries between physical, digital, and biological realms, between natural and artificial, between physical and virtual worlds, between online and offline activities, and between reality and virtuality.

⁴ Chris Holder, *et al.*, “Robotics and Law: Key Legal and Regulatory Implications of the Robotics Age (Part I of II),” *Computer Law & Security Review* 32, no. 3 (June 2016) // <https://doi.org/10.1016/j.clsr.2016.03.001>.

⁵ A. S. George and A. H. George, “Industrial Revolution 5.0: The Transformation of the Modern Manufacturing Process to Enable Man and Machine to Work Hand in Hand,” *Seybold Report* 15, no. 9 (2020).

⁶ World Economic Forum, “Values, Ethics and Innovation: Rethinking Technological Development in the Fourth Industrial Revolution” // <https://www.weforum.org/whitepapers/values-ethics-and-innovation-rethinking-technological-development-in-the-fourth-industrial-revolution/>.

and technological capabilities,⁷ can change the world profoundly and irrevocably. As these technologies are becoming increasingly embedded in our daily routine, we begin to notice that they are not just neutral objects or tools, but rather, mediators,⁸ which, indeed, shape, interpret, transform, and make meaning of human lives.⁹ This technological development can also be characterised by participatory patterns in the economy and in other social spheres, including law. Social networks (Facebook, Twitter, Instagram), the sharing economy (Uber, Airbnb, Task Rabbit), and cryptocurrencies (Bitcoin, Ethereum) are just a few examples of this phenomenon.¹⁰

Most current legal instruments are arguably too slow and ineffective for regulating a transformative society. According to Prof. Milleire Hildebrandt, modern law grew up in a culture of print and remains firmly wedded to the medium of text. However, legal concepts and instruments, which were developed in a print culture, may be ineffective in the internet age.¹¹ According to Alexander Bard and Jan Soderqvist, language as information technology is the most important instrument of power balance in society. The internet has created the foundations for a new paradigm. It “has increased the quantity of information available in society to a revolutionary extent, which in turn has created entirely new social hierarchies”¹². In terms of the impact of language technology on societal development, Bard and Soderqvist argue that feudalism started with the emergence of written language, which was invented and started to develop more than 5,000 years ago. “It was at this time that the world was also blessed with the first formalized world religions—for example, Zoroastrianism, Brahmanism, Buddhism, Jainism, Taoism, and Judaism—and the first documented philosophy”¹³. The authors further summarise the era of capitalism, which came at the golden age of printed and mass-distributed ideologies. Capital was “set in motion after the breakthrough of the banknote press in the 17th century. With the arrival of capital, for the first time in history the symbol became even more important than what it claimed to represent”¹⁴. However, when the internet arrived with full force in the late 1980s and early 1990s, that society was endowed with an environment in which holism and generalism were fostered at the expense of the academic world’s atomism and specialism. The informationalist

⁷ David C. Vladeck, “Machines Without Principals: Liability Rules and Artificial Intelligence,” *Washington Law Review* 89, no. 1 (March 2014): 118.

⁸ Mireille Hildebrandt, *supra* note 1, 174.

⁹ Katinka Waelbers and Tsjalling Swierstra. “The Family of the Future: How Technologies Can Lead to Moral Change”; in: Jeroen van den Hoven, *et al.*, eds., *Responsible Innovation 1* (Dordrecht: Springer Netherlands, 2014) // https://doi.org/10.1007/978-94-017-8956-1_12.

¹⁰ Julija Kiršienė, Christopher Kelley, Deividas Kiršys, and Juras Žymančius. “Rethinking the Implications of Transformative Economic Innovations: Mapping Challenges of Private Law,” *Baltic Journal of Law & Politics* 12, no. 2 (2019).

¹¹ Mireille Hildebrandt, *supra* note 1, 140–218.

¹² Alexander Bard and Jan Soderqvist, *Syntheism: Creating God in the Internet Age* (Stockholm: Stockholm Text, 2014), 59.

¹³ *Ibid.*, 65.

¹⁴ *Ibid.*, 68.

paradigm was characterised by a narrative of the developing information, communication, and network society, by “interactivity as the dominant form of communication, by the cyber world as the geographical arena, and by attention rather than capital”¹⁵ as the driving force of the economy.¹⁶

As long as futurists, artists, and academics are modeling the scenarios for the impact of these industry revolutions on society, culture, politics, or nature, it is clear that businesses and even states can gain an advantage if they are prepared to benefit from these changes. If the law is NOT technologically neutral and is sensitive to its technological mediators, one might ask how the institution of law will change in this new paradigm of society; in other words, how are smart technologies able to change legal processes? For example: administration of processes, research and argumentation, analysis and evaluation of evidence and facts, decision-making; consultations, sharing of knowledge and experience, and learning. One might also ask whether traditional hierarchical regulatory models with a large scope may be replaced by decentralised, soft, inclusive governance tools as well as foresight instruments.

In a time of exponential technological change, companies spring up and die off with startling rapidity, and entire industries are being dramatically transformed. The law has long been unaffected by these winds of change. Even if the law is one of the slowest industries to adopt new technologies and legal culture in general is risk-averse, the impact of COVID-19 has been a catalyst for the further use of technology in the legal sphere, forcing legal system institutions toward digital transformation and adopting technologies that legal firms and courts should have been using long ago.

Susskind affirmed that technology can affect legal processes in two broad ways, corresponding closely to two types of changes. On the one hand, smart technologies can improve, refine, streamline, optimise, and turbo-charge our traditional ways of working. This is what judges and lawyers usually mean in this regard. They reflect on routine, repetitive, and often antiquated tasks and activities in their legal institutions, like courts or notary offices, and imagine that some set of systems could be introduced to bring new efficiencies and make life easier. This first broad approach to legal domain technologies is a form of process improvement called automation. It involves grafting new technology onto old working practices. It has dominated the theory and practice of LegalTech over the last 50 years or so. On the other hand, technology can play a very different role. It can displace and revolutionise conventional working habits and bring about radical change—doing new things rather

¹⁵ *Ibid.*, 70.

¹⁶ *Ibid.*, 58–79.

than old things in new ways. This is transformation. It is about using technology to allow us to perform tasks and deliver services that would not have been possible, or even conceivable, in the past. The impact of transformative technologies can be profound. Transformation brings disruption, whereas automation sustains traditional ways of working. However, while automation is reassuringly familiar, the dominant general trend in the world of technology is very clearly towards transformation¹⁷.

However, Susskind did not consider the importance of innovations in the legal domain.¹⁸ Transformation and innovation are buzzwords in the current technological and business landscape. According to Merriam-Webster, transformation means “an act, process, or instance of transforming or being transformed,”¹⁹ while innovation is “the introduction of something new.”²⁰ Innovation and transformation are often used synonymously, but they actually have different meanings. In some instances, transformation can lead to innovation. In others, innovation can lead to transformation, but this mutually causal relationship is often overlooked because of misinterpretations. Schumpeter, who showed that innovation is the source of economic fluctuations, noted that “the pure idea is not adequate by itself to lead to implementation.”²¹ Innovation is one of the drivers of digital transformation (change) and usually, but not necessarily, precedes the change process. Systemic innovation is different than autonomous innovation in the digital domain and requires changes in the supply chain, management, information technology (digital innovation), learning processes, etc. Systemic innovation depends on a series of interdependent innovations; hence, in most cases, it is a trigger for a business-wide transformational process. Transformation describes processes that persist after innovation enters the scene. The transformation of the legal sphere is different than in other domains or industries, as it is not enough simply to meet changing market demands, but also to address public interests such as, for example, improving access to justice. So, the transformation of the legal domain is usually correlated with systematic innovations.

Figure 1 presents some of the areas in which AI technologies have started to improve human efficiency in the legal domain:

¹⁷ Richard Susskind, *The Future of Law: Facing the Challenges of Information Technologies* (Oxford: Clarendon Press, 1998).

¹⁸ *Ibid.*

¹⁹ See <https://www.merriam-webster.com/dictionary/transformation>.

²⁰ See <https://www.merriam-webster.com/dictionary/innovation>.

²¹ J. A. Schumpeter, *The Theory of Economic Development* (New Brunswick, NJ: Transaction Publishers, 1983).

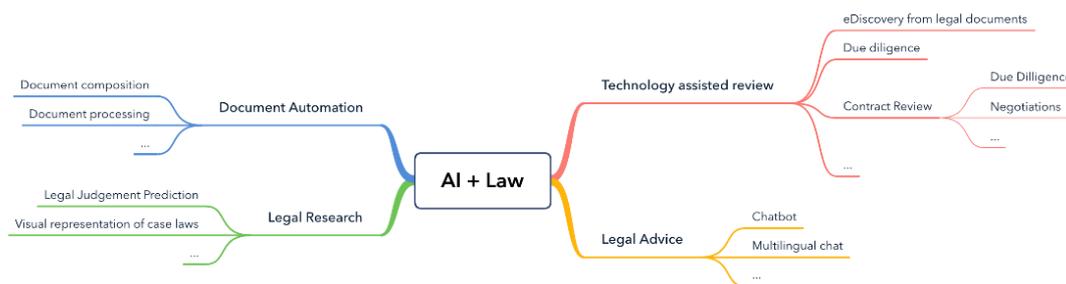


Figure 1: AI technologies in the Legal Domain

- *Document Automation.* Legal work traditionally requires a lot of paperwork, and hence a significant portion of the workload of legal professionals involves composing and processing documents. AI can ingest several such documents and learn to generate similar documents for humans. The AI can also learn to extract useful information from these documents and execute predefined logic.

- *Legal Research.* Identifying the cases applicable to current matters can profoundly impact the outcome of the case. The AI can improve the case law search and enable it to recommend more relevant cases. The ability of computers to “understand” judgements can also provide insights into the likely outcome of the case and, in some cases, help to present cases in a manner that is better for analysis.

- *Technology-Assisted Review.* AI can hugely reduce the human burden of due diligence on existing legal documents and extract critical information from a large volume of documents in highly complex matters. AI-assisted review would enable legal professionals to do more accurate due diligence and enable them to process larger volumes of legal documents.

- *Legal Advice.* Traditionally, lawyers have been the primary source of legal advice. With the rise of internet blogs and forums, people have another avenue to get legal advice. With the recent advancements in HLT and AI in the form of conversational AI, the possibility of machines processing legal queries and providing accurate legal advice has become a reality. When added to machine translation, providing legal advice in multiple languages has become a possibility.

- *E-Filing.* This is the electronic submission of information that is required by law. It is up to the regulating agency to set requirements. An e-filing system is a complete end-to-end solution for the online filing of complaints, written statements, replies, and various applications related to cases. Both civil and criminal cases can be filed before any court or other legal institution. An e-filing system can be designed in multilingual environments to reach a wider scope of litigants. E-filing systems provide several benefits: (a) they save time, money, and travel for lawyers and

clients; (b) they obviate the need to visit the court physically; (c) they reduce the need for meetings between clients and lawyers; (d) they allow for the automatic digitisation of case records; (e) they have positive impact on the environment by reducing paper use; and (f) they allow for reduced file processing time and immediate document availability to the parties and the court. Because of COVID-19, many courts are implementing e-filing or consider it a real possibility for improving services and, accordingly, access to justice.

- *Outcome prediction* in a legal case has always been crucial to the practice of law. A reasonable assessment of the potential legal consequences is important for legal counselling and judgment delivery. As compared to the traditional tools used by lawyers and judges for prediction (professional experience, empirical information, etc.), AI offers a much more sophisticated approach. Predictive analytics using AI tools and technologies (like NLP and machine learning) has made predictive justice possible. Predictive justice refers to using analyses of large amounts of data by means of AI-enabled technologies to predict the outcomes of legal disputes. In AI jargon, the term "predictive" is linked to the possibility of predicting future results through inductive analysis, which identifies correlations between input data and output data. AI tools have more precision as they can process a vast set of information, which is humanly not possible. Humans working in tandem with AI can outperform either working in isolation. AI models with predictive analytics can offer solutions to a number of problems faced by judicial systems, such as long pendency, legal inconsistency in the application of law, etc. For example, a dispute filed before a court where a violation seems likely (as concluded by a predictive AI) could be prioritised. Cases that involve repetitive or simple applications of law can be automated. Judges can thus focus on more complicated cases and increase the efficiency of the court system. Parties to the dispute may also decide to take the case outside the court in situations where success is unlikely (as predicted by AI) and to avoid wasting time and resources in long litigation. This might also help to reduce the information asymmetry between the parties. The main problem with prediction is that AI technologies project a past version of their human counterparts into the future through adaptive learning techniques.

So, document automation, technology-assisted review, and e-filing can be considered fields of *automation* with some portion of innovation. However, legal advice and especially legal research (outcomes prediction, predictive justice, etc.) are fields in which *innovations* and *transformation* are necessary. However, to realise the full potential of these technologies, different forms of AI must be applied to varied legal functions. This inference brings us back to the main question of whether technology can be leveraged to improve access to justice.

2. ACCESS TO JUSTICE IN THE POST-PANDEMIC WORLD

Every right can be effectively enjoyed only if persons have the opportunity to claim it and to protect it through legal means. This is an important element in upholding the rule of law. The right of access to justice is enshrined in all major international human rights documents—the International Covenant on Social, Political, and Cultural Rights, the European Convention on Human Rights, the African Charter on Human and Peoples' Rights, and others—as well as being widely included in national constitutions. Although the term “access to justice” is not used in these documents, the doctrine draws it from several related rights, such as the right to a fair trial, due process, etc. Therefore, the meaning of this right to access justice can be narrowly interpreted as the possibility for individuals to bring claims before a court and have a court adjudicate them. In a broader sense, it contains “the right of an individual not only to enter a court of law, but [also] to have his or her case heard and adjudicated in accordance with substantive standards of fairness and justice.”²²

As research shows, for about two-thirds of the world's population, it is not feasible to exercise legal rights.²³ Most such people are members of socially vulnerable groups, such as women, older people, minorities, people with disabilities, refugees, people without sufficient education or language skills, or people with special needs.

The first sentence of the preamble to the Universal Declaration of Human Rights states that “recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world.”²⁴ Unfortunately, too often, one of the key elements of inherent human rights—justice—is not accessible. One reason is the lack of access to a wide range of legal services. For example, during the pandemic, lockdowns and restrictions on movement exacerbated the problem of digital exclusion and made it difficult for some societal groups to exercise their rights to information, health care, education, family relationships, privacy, and other rights. Refugees, persons in forced isolation due to the pandemic, residents who do not understand the local language, tourists and foreign citizens, older people, and persons who lack special skills and legal capacity faced various constraints and difficulties because of difficult or insufficient access to justice.

²² Francesco Francioni, ed., *Access to Justice as a Human Right* (Oxford: Oxford University Press, 2007), 64.

²³ World Justice Project, “Global Insights on Access to Justice 2019” // <https://worldjusticeproject.org/our-work/research-and-data/global-insights-access-justice-2019>; United Nations, “World Social Report 2020: Inequality in a Rapidly Changing World” // <https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/01/World-Social-Report-2020-FullReport.pdf>.

²⁴ *Universal Declaration of Human Rights (UDHR)*, United Nations General Assembly, New York, 1948.

To analyse how the COVID-19 pandemic has influenced access to justice and how AI and other technological means can help to improve it, it is first necessary to outline the elements of the term “access to justice.” Since outlining all the elements of this right would go beyond the constraints of this article and it would be virtually impossible to provide an exhaustive list of such elements,²⁵ only the procedural aspect of access to justice is examined, as it is most likely to be influenced by technological advancements.

Cost of litigation. Access to justice cannot be achieved if the cost of defending your rights in court is so high that you are precluded from seeking justice for economic reasons. The cost of litigation includes both the costs of legal representation and the costs of submitting a claim to the court (court fees), participating in court hearings, calling witnesses and experts to testify in one’s case, and using translators.²⁶

The measures that states have adopted in light of the pandemic (social distancing, stay-at-home orders, travel restrictions) have facilitated a worldwide move from regular court proceedings to online hearings. Developing technological instruments that make such online proceedings secure and accessible to all parties could have a positive impact on the costs of litigation by minimising travel fees to the hearing and by excluding the fees that in some cases have to be paid to the court to secure the participation of witnesses.²⁷

Transparency of the proceedings. One element of access to justice is ensuring that the case is heard in public. The introduction of online court proceedings during the COVID-19 pandemic and continuing to use such online proceedings after the pandemic also raises the issue of ensuring that courts stay open to the public.

Equality of arms. Under this principle, each party to a legal dispute must have an approximately equal chance to participate in the hearing, present statements, and provide evidence and witnesses. When transitioning to online court proceedings or using other technology-based instruments to bring a claim or participate in the court process, it is important to consider inequality in access to online services and a lack of technological literacy. The pandemic has revealed the extent of this technological inequality, which in Lithuania became most prominently evident in the education sector when online learning was introduced, and it became apparent that a lack of electronic devices or access to the internet was a hurdle for some schoolchildren to

²⁵ Julia Bass, W. A. Bogart, and Frederick H. Zemans, eds., *Access to Justice for a New Century—The Way Forward* (Toronto: Law Society of Upper Canada, 2005), 3.

²⁶ European Parliamentary Research Service, *Common Minimum Standards of Civil Procedure. European Added Value Assessment*, Annex I (Strasbourg: European Parliament, 2016), 39.

²⁷ For example, under the Code of Civil Procedure of the Republic of Lithuania, Art. 90, the party on whose request the witness is called to testify in the court proceedings has to make an advance payment into the court’s bank account to finance the travel costs of that witness. If the party does not pay the amount, the court can refuse to hear the witness’s testimony.

participate in online learning. A lack of technological means can also have a greater impact on people who are already in vulnerable economic and social positions, thus making it even harder for such individuals to access legal services.²⁸ Some research suggests that even prior to the pandemic, socioeconomically disadvantaged people were more vulnerable to experiencing legal problems.²⁹

According to the Access to Justice Scorecard 2020 survey by the Queensland Law Society, the vast majority of respondents agreed that technology has an important role in access to justice, especially online access to information about laws and legal rights, the capacity to access and upload documents online, and video/teleconferencing of hearings.³⁰ However, few have indicated that the use of AI has facilitated better access to justice. So, the main question is not whether imminent technological advances may bring benefits to clients in the form of lower prices and greater ease of use of legal services, but whether technology can be leveraged to improve access to justice. In other words, even if customer experience has been considerably improved using AI in other industries—from banking and retail to consumer electronics and transportation—can it also enhance access to legal services?

3. AI TECHNOLOGIES IN THE LEGAL DOMAIN: IMPLICATIONS FOR LEGAL SERVICES

At close to \$1 trillion globally, the legal services market is one of the largest in the world, but it remains profoundly under digitised.³¹ The legal profession is fairly conventional in its approach and hesitant to adopt new technologies. Each breakthrough in the field of law has been revolutionary.³² Legal AI mainly focuses on applying AI technology to help with legal tasks. Many researchers have devoted considerable efforts over the past few decades to promote the development of LegalAI,³³ and their efforts have led to tremendous advances.³⁴ LegalAI can

²⁸ OECD and Law & Justice Foundation of New South Wales, *Access to Justice and the COVID-19 Pandemic: Compendium of Country Practices* (Paris and Sydney: OECD/Law & Justice Foundation of New South Wales, 2020), 13–14 // [http://www.lawfoundation.net.au/ljf/site/articleIDs/5BE8DA8F5FA5EF7E852586B100025DB7/\\$file/oecd-access-to-justice-covid19.pdf](http://www.lawfoundation.net.au/ljf/site/articleIDs/5BE8DA8F5FA5EF7E852586B100025DB7/$file/oecd-access-to-justice-covid19.pdf).

²⁹ *Ibid.*, 2.

³⁰ For example, on 24 March 2020, the Supreme Court of the United Kingdom conducted its first ever remote hearing. The Dubai Courts confirmed that from 19 April 2020, all hearings would be conducted on Microsoft Teams.

³¹ See <https://legalpediaonline.com/artificial-intelligence-law/>.

³² From typewriters to computers, from using libraries solely for legal study to using internet portals to find relevant case laws, from fax machines to emails.

³³ Artificial general intelligence is a theoretical application of generalised AI to any domain, solving any problem that requires AI. Artificial narrow intelligence, sometimes known as “weak AI,” refers to any AI that can outperform a human in a narrowly defined and structured task. In this paper, we define legal technologies tools using weak AI as LegalAI.

³⁴ E.g., Ilias Chalkidis *et al.*, “Neural Legal Judgment Prediction in English” (2019) // <https://arxiv.org/pdf/1906.02059.pdf>; Yadong Cui, *Artificial Intelligence and Judicial Modernization*

theoretically play a significant role in the legal domain, as it can reduce heavy and redundant work for legal professionals, and it can also provide a reliable reference to those who are not familiar with the legal domain, serving as an affordable form of legal aid. Having said this, we agree with the opinion of the European Ethical Charter on the use of AI in judicial systems and their environment³⁵ that we must differentiate between commercial discourse about benefits of LegalAI and the reality of the use and deployment of these technologies. Due to the non-technical nature of this article, we do not pretend to present a full review and evaluation of the actual state of LegalAI. As stated by sitting Judge of the Supreme Court of India Justice D. Y. Chandrachud, technology is relevant when it fosters efficiency, transparency, and objectivity, because its primary purpose is to give common people greater access to justice in the long run.³⁶

According to Susskind, the superficially straightforward question “Can technology be leveraged to improve access to justice?” conceals at least five questions³⁷. The first question is whether it is *technically possible* for machines to replace legal professionals. The second question poses the issue of whether, even if it were technologically possible, it would be *morally acceptable* for machines to take on any judicial functions. The third question concerns whether such systems would be *commercially viable*, i.e., whether their economic benefits would outweigh their costs? The fourth question is whether this would be *culturally sustainable*; in other words, whether such systems would be assimilated without rejection into legal institutions dominated by age-old procedures with human judges and other law professionals at their core. Finally, there is a legal question of whether it is *jurisprudentially coherent* to develop such instruments, i.e., whether there is anything specific about the structure and nature of judicial decision-making itself that places it, partly or entirely, beyond the scope of computation. In this paper, we elaborate on at least three of these questions, namely technical feasibility, legal acceptability, and moral acceptability, leaving the questions about commercial and cultural viability for further research.³⁸

(Cham: Springer/Shanghai People’s Publishing House. 2020); Kevin D. Ashley, *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age* (Cambridge: Cambridge University Press, 2017).

³⁵ *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*, European Commission for the Efficiency of Justice, Brussels, 2018 // <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

³⁶ Sanjana Shikhar, “The Role of Artificial Intelligence in Law” (2021) // <https://blog.iplayers.in/role-of-artificial-intelligence-in-law/>.

³⁷ Richard Susskind, *supra* note 17.

³⁸ *Ibid.*

3.1. CHALLENGES FOR AI-BASED TECHNOLOGIES IN THE LEGAL DOMAIN

When answering the question of whether it is technically possible for machines to be used in the legal domain, the issues of legal language vs. common language, data and algorithms must be considered. Machine learning and its subfield, deep learning, are often considered branches of AI since a well-performing algorithm may produce automated results that appear “intelligent.” If performing well, machine and deep learning algorithms may produce automated results that approximate those a similarly situated human being would have made. However, existing AI tends to involve “narrow” intelligence. The power and practicality of AI are largely undisputed. However, one of its longest-standing difficulties continues to be human language.

A long-standing goal of AI has been to build intelligent agents that can function with the linguistic dexterity of people, which involves such diverse capabilities as participating in a fluent conversation, using dialogue to support task-oriented collaboration, and engaging in lifelong learning through processing speech and text. There has been much debate about whether this goal is, in principle, achievable since its component problems are arguably more complex than those involved in space exploration, mapping the human genome, or the classification of cats and dogs in pictures. In fact, enabling machines to emulate human-level language proficiency is well understood as an AI-complete problem³⁹—one whose full solution requires solving the problem of AI in general. For the uninitiated, the complexities of natural language are not self-evident; after all, people seem to process language effortlessly. But the fact that human language abilities are often taken for granted does not make them any less spectacular. When analysed, the complexity of the human language facility is, in fact, staggering—which makes modelling it in silico a very difficult task indeed. This paper does not consider the complexity of the issues that make language difficult for an AI agent. As a brief illustration of this complexity, we will offer the example of ambiguity. There are many types of ambiguity in natural language. A few examples are morphological ambiguity, lexical ambiguity, syntactic ambiguity, semantic ambiguity, and pragmatic (context) ambiguity.⁴⁰ In short, the statistical/machine learning approach does not attempt to compute meaning. Instead, practitioners proceed as if words were a sufficient proxy for their meanings, which they are not. In fact, the words in a sentence are only the tip of the iceberg when it comes to the full, contextual meaning of sentences. Confusing words with their meanings is a fraudulent approach that could lead to disastrous results.

³⁹ Alan Turing, “Computing Machinery and Intelligence,” *Mind* 59, no. 236 (1950).

⁴⁰ M. McShane and S. Nirenburg, *Linguistics for the Age of AI* (Cambridge, MA: MIT Press, 2021).

In LegalAI commercial discourse, considerations about the relationship between common (plain) language and legal domain language are often omitted. Legal language differs significantly from common language. Legal English has already been a subject of scientific research for a few decades.⁴¹ However, Lithuanian, like many other smaller languages, is still under-researched. David Mellinkoff defined legal language as “the customary language used by lawyers in those common law jurisdictions where English is the official language. It includes distinctive words, meaning, phrases and modes of expression ... legal language is a label for a speech pattern with a separate identity.”⁴² The law is written, and this, of course, provides a clear advantage from a technical point of view. In other words, language is intimately involved in the process of the application of the law. This leads to the question of whether everyone knows the law, since the law is almost entirely written down in accessible statutes, case reports, textbooks, and commentaries. However, the Statute Law Society in a submission to the Renton Committee described legal language as “legalistic, often obscure and [involving] circumlocutions, requiring a certain type of expertise in order to gauge its meaning. Sentences are long and involved, the grammar is obscure, and archaisms, legally meaningless words and phrases, [and] tortuous language ... abound.”⁴³ The law’s effectiveness as a system of justice depends on its legitimacy, which in turn demands that the presumption of knowledge be tenable as a description of the state of affairs to which it refers; the remoteness of the language of the law from ordinary speech argues against the presumption. Thus, it is no surprise that the Development Program of the Justice System in Lithuania (2021–2030) has indicated several causes of heavy burdens on the courts. A few of them are as follows: “The number of consumer requests and complaints is steadily increasing,” “Courts deal with simple cases where there is no legal litigation”, and “Lack of public legal knowledge in defending their violated rights.”⁴⁴ Unfortunately, the ordinary person rightly believes himself or herself to be its addressee. Of course, the defenders of legal language could say that it is a well-adapted discourse to its function, even though the distinctive characteristics

⁴¹ Just to mention a few examples: David Mellinkoff, *The Language of the Law* (New York: Little, Brown and Co., 1963); Peter Goodrich, *Legal Discourse* (New York: Palgrave Macmillan, 1987); Alfred Philips, *Lawyers’ Language: How and Why Legal Language is Different* (New York: Routledge, 2003); Michal Freeman and Fiona Smith, eds., *Law and Language*, Vol. 15 (Oxford: Oxford University Press, 2011).

⁴² David Mellinkoff, *supra* note 41, 3.

⁴³ Usama Mubarak, *Role of Legislative Drafting in Victimization of Citizens* (2020) // <https://ssrn.com/abstract=3656529> or <http://dx.doi.org/10.2139/ssrn.3656529>.

⁴⁴ *Development Program of the Justice System in Lithuania (2021-2030)*, Government of the Republic of Lithuania // <https://tm.lrv.lt/uploads/tm/documents/files/dokumentai/Administracin%C4%97%20informacija/PI%C4%97tros%20programos/2021-2030%20metu%20teisingumo%20sistemas%20pletros%20programa.pdf>.

distancing it from common speech have resulted in the creation of a preserve restricted to competent users.⁴⁵

Other technical topics concern algorithms and data. Since most legal resources and data are presented in a text form, such as judgement documents, contracts, and legal opinions, they are suitable for HLT. Like machine learning or deep learning, HLT is a subset of AI. It is the branch of AI that enables computers to analyse, interpret, and manipulate human language. In its broadest sense, HLT refers to any work involving the computational processing of natural language. HLT, especially its subfield—natural language processing (NLP)—has its roots in linguistics, where it emerged to enable computers to process natural language literally, and the term “NLP” encompasses all methods of automating the processing of natural language. However, over the past few decades, NLP has taken on the strong default connotation of involving knowledge-lean (essentially, semantics-free) machine learning and deep learning methodologies to ingest and process large unstructured speech and text datasets effectively. Therefore, in the historically recent and current context, there is a juxtaposition between NLP and the second subfield of HLT—natural language understanding (NLU). Figure 2 represents the development of NLP over time, its relation to the amount of labelled data, and its explainability. The green line shows the direction of state-of-the-art development, which tends toward the full autonomy of AI systems.

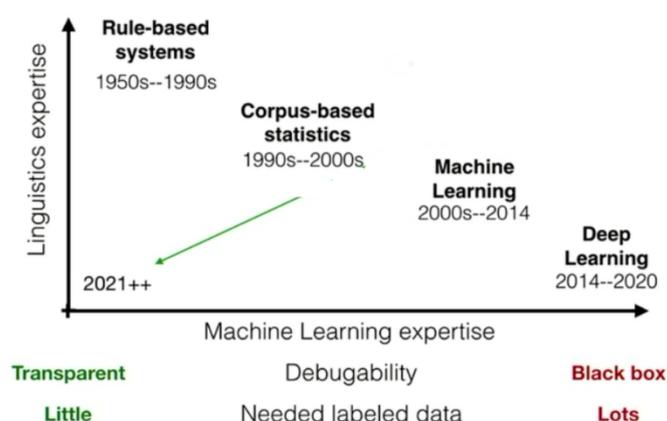


Figure 2: The Development of NLP

There are no suitable NLP tools and systems for processing common language or any language domain. However, as we said above, legal domain language is a very

⁴⁵ This critique becomes sharper and more radical in theories that suggest that lawyers foster this disease themselves in the pursuit of their professional interests.

special case.⁴⁶ There is insufficient space in this paper to explore all of the issues related to NLP of the legal domain language. We therefore provide two examples: insufficient training data and context limitations, which are among the biggest obstacles to adopting state-of-the-art HLT in the legal domain.

To process language data better, be it text or sound, we need to develop systems that have a memory element to provide context. Most advanced transformer-based processing algorithms with attention mechanisms do well with context but perform poorly with legal documents because they cannot process long sequences. Their input is limited to 512 tokens.⁴⁷ The average legal document is much longer. It must be split into subparts for processing. In this way, the context of the document is lost. The recently introduced longformer⁴⁸ offers a very promising approach that scales linearly with sequence length, making it easy to process documents of thousands of tokens or longer. However, this is still not enough to analyse legal documents, and so far, the longformer has only been tested on English and Chinese texts.⁴⁹

In addition, to build efficient AI algorithms, they must be trained on really large amounts of data. Historical data constitute the basis for training and for predictive analytics. In the legal domain, this constitutes a major challenge, especially for under-resourced non-English languages (such as Lithuanian). In legal terms, AI platforms require a vast amount of historical data to develop a level of experience and begin learning how to make certain decisions. There are at least some representative open-access legal text datasets in English.⁵⁰ In Lithuanian, there are very few.⁵¹ Legal documents usually contain extensive personal data and confidential information. Deep learning algorithms do not generalise well from only a few examples (as the human brain can). For this reason, we have an unfair language-based situation because small languages (such as Lithuanian) became under-resourced and lagged behind English.

⁴⁶ Some scholars have suggested that disposing of legal jargon and using plain English could mitigate many of the difficulties. See, e.g., Rabeeah Assy, *Injustice in Person: The Right to Self-Representation* (Oxford: Oxford Scholarship Online, 2015); John W. Hager, "Let's Simplify Legal Language," *Rocky Mountain Law Review* 32 (1959).

⁴⁷ A token is a technical term for the smallest structural piece of text. In the case of the Transformers tokenizer, it may contain a word or only part of a word (e.g., one word may form two or three tokens).

⁴⁸ I. Beltagy, "Longformer: The Long-Document Transformer" (2020) // <https://arxiv.org/pdf/2004.05150.pdf>.

⁴⁹ Chaojun Xiao, *et al.*, "Lawformer: A Pre-Trained Language Model for Chinese Legal Long Documents" (2021) // <https://arxiv.org/pdf/2105.03887.pdf>.

⁵⁰ See, e.g., <https://legal-linguistics.net/data-collections/>; <https://archive.org/details/ECHR-ACL2019/>; <https://archive.org/details/Law2Vec/>; <https://www.atticusprojectai.org/> (CUAD—a corpus of 13,000+ labels in 510 commercial legal contracts with rich expert annotations).

⁵¹ Most significant is the widely used corpus of European Parliament legal texts: <https://www.statmt.org/europarl/>. Anonymized Lithuanian court documents are available in open access on <http://liteko.teismai.lt/viesasprendimupaieska/detalipaieska.aspx?detali=2> but not as a digital resource suitable for machine learning.

On the other hand, there are two main approaches to building AI systems: the data-driven approach and the model-driven approach. The model-driven approach is to build new models and new algorithmic manipulations to improve performance. The data-driven approach may improve data quality and/or data governance to improve the performance of a specific problem statement. The data-driven approach is a new way of thinking, enabled by machine learning, that permits machines to spot connections and correlations that humans may not even know or suspect. One important aspect of AI projects is that bad data or a lack of data (including clean, standardised, and labelled data) can never lead to great performance, however good a model is. Also, ethics must be taken seriously because data can contain biases that can be transferred into the AI model during the training process and can cause biased decisions from the AI system. Unfortunately, there are still issues with AI transparency, the so-called “black box problem”. Powerful AI machine learning models, in particular deep neural networks, tend to be very hard to explain. Sometimes there is the dilemma of having to accept a particular model that works much better than another, although it is more difficult to understand and explain how excellent results/predictions are obtained. This problem is closely related to another major problem regarding bias in the data used to train AI/machine learning algorithms. That bias will be present in the decisions that the algorithms make. This is clearly unacceptable, but it is difficult to identify in final decisions and to explain. Facebook’s machine learning tools predict users’ preferences better than any psychologist. AlphaFold,⁵² a program built by DeepMind, has produced the most accurate predictions yet of protein structures based on the amino acids they contain. Both are completely silent on why they work, e.g., why they prefer this or that information. For now, this is unacceptable for humans because they remain deeply uncomfortable with such a situation. Humans do not like dealing with a black box—they want to know *why*.

Thus, AI algorithms still tend to work poorly, or not at all, in areas that are conceptual, abstract, value-laden, open-ended, policy- or judgment-oriented, or that require common sense or intuition, involve persuasion or arbitrary conversation, or involve engagement with the meaning of real-world humanistic concepts, such as societal norms, social constructs, or social institutions.⁵³ Hinton has expressed the opinion that state-of-the-art deep learning technologies will be able to do everything if only quite a few conceptual breakthroughs take place.⁵⁴ To sum all this up,

⁵² See <https://www.deepmind.com/research/highlighted-research/alphafold>.

⁵³ See, e.g., Garry Marcus, “Deep Learning: A Critical Appraisal” (2018) // <https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf>.

⁵⁴ See <https://www.technologyreview.com/2020/11/03/1011616/ai-godfather-geoffrey-hinton-deep-learning-will-do-everything/>.

conceptual breakthroughs in NLP must take place to apply LegalAI fully in the legal domain. This requires multidisciplinary research efforts.

3.2. MORAL AND LEGAL ASPECTS OF AI IN THE LEGAL DOMAIN

Sussking's second question reflects *moral acceptability*. The judicial system operates on trust. Without it, there is no avenue for self-represented litigants to access justice. The same is true for AI technologies in the legal system. Like the trustworthy adoption of drugs or surgical procedures, there must be evidence that AI is effective at achieving a specified objective without undue risk. Trustworthy adoption of AI means adoption based on sound evidence of the extent to which AI-enabled processes in the judicial and other legal domains help to achieve justice. Court users must trust that AI and judicial agents produce desirable outcomes and that they can be held accountable if they do not. This requires an operational definition of desirable outcomes and the ability to measure the extent to which these are met. If humans can achieve trustworthy technological adoption—or, equally importantly, avoidance of adoption—in medicine and aviation, they can do so for AI in the legal domain.⁵⁵ In other words, AI technology needs to prove its trustworthiness to beneficiaries of the legal system. On the other hand, the use of AI in the judiciary system can also instil public trust and thereby help to scale AI-based technologies in other public services like health, education, finance, etc.

In terms of moral acceptability, the aspects of *an ethical framework in deploying AI in the administration of justice* and *AI trustworthiness to users of the justice system* are very important. An ethical framework is a precursor to the development and deployment of AI. The European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment published by CEPEJ provides five broad principles: respect for fundamental rights; non-discrimination; quality and security; transparency, impartiality and fairness; and user control. The last principle stands out as it assists in steering clear of the dehumanisation process of the justice system. It precludes a prescriptive approach of the use of AI and preserves the autonomy of the judge—to review the judicial decision and the data used to produce the result proffered by AI. We agree with P. Sales that “coding will reflect the unspoken biases of the human coders,” that coding algorithms are closed systems that “may not capture everything of potential significance for the resolution of a human problem,” and that “the open-textured nature of ideas like justice and fairness

⁵⁵ V. Li and S. la Roque-Doherty, “Toward Smarter Courts: Artificial Intelligence Has Made Great Inroads—But Not as Far as Increasing Access to Civil Justice,” *ABA Journal* 107, no. 2 (April/May 2021) // <https://go.gale.com/ps/i.do?p=EAIM&u=anon~b1a1800&id=GALE|A655942226&v=2.1&it=r&sid=bookmark-EAIM&asid=e697e6f6>.

creates the possibility for immanent critique of rules being applied and leaves room for wider values, not explicitly encapsulated in law's algorithm, to enter the equation leading to a final outcome."⁵⁶ We must note that the goal of LegalAI is not to replace legal professionals but to help them in their work. As a result, we should regard the results of the models only as a reference. Otherwise, the legal system will no longer be reliable. For example, professionals can spend more time on complex cases and leave the simple ones to the model. However, for safety, these simple cases must still be reviewed. In general, LegalAI should play a supporting role to help the legal system.⁵⁷

A recent proposal of the European Commission for the regulation of AI⁵⁸ describes AI as a fast-evolving family of technologies that can bring a wide array of economic and societal benefits. The main objective of this proposal is to ensure the free circulation of goods and services embedding AI by avoiding further fragmentation and a potentially contradictory national legal framework. As the proposal mentions, the definition of AI, as well as its main characteristics, have been widely discussed in research.⁵⁹ So, if the notion of an AI system were clearly defined, at the same time as providing the flexibility to accommodate future technological developments, it would be very beneficial for legal certainty. The definition could be based on such functional characteristics of the software as the ability to provide predictions, recommendations, or decisions that influence the environment with which the system interacts in physical or digital dimensions with varying levels of autonomy. Thus, AI-based technologies should rely on large and varied datasets that may be embedded in any product or service.

It is noteworthy that "aside from the many beneficial uses of artificial intelligence, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices."⁶⁰ Also, the problems of opacity, complexity, bias, a certain degree of unpredictability, and the partially autonomous behaviour of certain AI systems should be addressed.⁶¹ Therefore, the

⁵⁶ See <https://judicature.duke.edu/articles/algorithms-artificial-intelligence-and-the-law/>.

⁵⁷ Haoxi Zhong, "How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence" (2020) // <https://arxiv.org/abs/2004.12158>.

⁵⁸ *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, European Commission, COM(2021) 206 final, 2021/0106, Brussels, 2021.

⁵⁹ Julija Kiršienė, Edita Gruodytė, and Darius Amilevičius, "From Computerised Thing to Digital Being: Mission (Im) Possible?" *AI & SOCIETY* 36, no. 2 (2021).

⁶⁰ *Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, European Commission, Brussels, 2021.

⁶¹ Darius Amilevičius, "Machine Bias and Fundamental Rights"; in: John Gordon, ed., *Smart Technologies and Fundamental Rights* (Leiden: Brill, 2020); *Resolution of 20 October 2020 on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies*, European Parliament, 2020/2012(INL), Strasbourg, 2020; *Presidency Conclusions—The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change*, Council of the European Union, 11481/20, Brussels, 2020.

mentioned proposal suggests setting specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons, for example, legal requirements in relation to data and data governance, documentation and recording keeping, transparency and the provision of information to users, human oversight, robustness, accuracy, and security. So, AI systems intended to be used in the legal sphere, in view of the nature of the activities in question and the risks relating thereto, should be considered high-risk.

However, as the proposal shows, not all technologies used in the legal sphere would be considered high-risk. For example, it is suggested that "AI systems specifically intended to be used for administrative proceedings by tax and customs authorities should not be considered high-risk AI systems," in contrast with high-risk analytical tools used by "enforcement authorities for individual risk assessments, polygraphs and similar tools or to detect the emotional state of natural person, to detect 'deep fakes', for the evaluation of the reliability of evidence in criminal proceedings, for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons, or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups, for profiling in the course of detection, investigation or prosecution of criminal offences, as well as for crime analytics regarding natural persons."⁶²

In other words, in interactions between law enforcement authorities and people, there is a power imbalance that may lead to surveillance, arrest, or deprivation of a natural person's liberty, as well as other fundamental rights.⁶³ It is therefore usually appropriate to classify AI technologies used in law enforcement procedures as high-risk because accuracy, reliability, and transparency are particularly important to avoid adverse impacts, retain public trust, and ensure accountability and effective redress.

4. CASE ANALYSIS: CHALLENGES OF DIGITALISATION OF LEGAL SERVICES IN PRACTICE

When it comes to integrating AI with court services for the public, Lithuania is behind many countries.⁶⁴ For example, as early as 2016, the high court in Hebei Province

⁶² Other high-risk technologies include systems used to evaluate the credit score or creditworthiness of natural persons.

⁶³ In particular, if the AI system is not trained with high-quality data, does not meet adequate requirements in terms of its accuracy or robustness, or is not properly designed and tested before being put on the market or otherwise put into service, it may single out people in a discriminatory or otherwise incorrect or unjust manner. Furthermore, the exercise of important procedural fundamental rights, such as the right to an effective remedy and to a fair trial, as well as the right of defence and the presumption of innocence, could be hampered, in particular, where such AI systems are not sufficiently transparent, explainable, and documented.

⁶⁴ OECD and Law & Justice Foundation of New South Wales, *supra* note 28; European Parliamentary Research Service, *Language Equality in the Digital Age* (Strasbourg: European Parliament, 2017) //

(China) introduced the smart court concept.⁶⁵ The smart court includes online hearings, electronic case filing, transcription of audio records (speech to text), case party identification, and automated document production, and it delivers related laws, regulations, and authoritative cases to judges (advanced and semantic search, data mining, and reasoning).

In 2021, our interdisciplinary team, consisting of law, IT, and linguistics researchers, finished our large Semantika-2 project.⁶⁶ The main scope of the project was to create a set of tools and solutions for document automation and technology-assisted review for the Lithuanian language. The project was funded by the European Union Structural funds and Lithuanian Government funds. For this reason, all the results are free of charge for public use. The need for this kind of project is dictated by the need to make breakthroughs in Lithuanian language technologies,⁶⁷ because the automation of processes and services that are related to human language mainly depends on them. As is commonly known, small and under-resourced languages lag behind the progress of the English language.⁶⁸

The main set of tools and solutions created during the development of Semantika-2 consists of automatic transcription of speech records into digital text (speech to text), automatic summarisation of long documents, semantic analysis tools (named entity recognition and extraction, aspect-based sentiment analysis⁶⁹), automatic spell checking, and linguistic analysis tools. All the solutions were realised according to modern cloud-ready architecture requirements. Where possible, AI and machine learning technologies were used to develop NLP solutions. But in some cases, this was not possible because some recent open-sourced algorithms are suitable for English but do not work well with morphologically rich languages, such as Lithuanian. In those cases, rule-based methods or hybrid methods were used.

As discussed in Part III, specific domain languages (legal, medical, administrative, etc.) differ from general language usage at least to some extent (vocabulary, semantics, patterns, syntactical structures, etc.). There has been

<https://op.europa.eu/en/publication-detail/-/publication/fa0a50e7-cda4-11e7-a5d5-01aa75ed71a1/language-en/format-PDF>.

⁶⁵ However, significant investments in information and communication technologies for e-justice services have caused some to question whether these investments have achieved the expected ends. Thus, how to evaluate e-justice services becomes an urgent theoretical and policy issue in the process of e-justice construction in China. E-justice value is not clearly defined in theory in China and is not easy to measure in practice.

⁶⁶ See www.semantika.lt. Acknowledgements to the researchers who worked on the Semantics project for the use of the research results.

⁶⁷ Statal Commission for Lithuanian Language, *Guidelines for the Development of the Lithuanian Language in Information Technology 2014–2020* (Vilnius: Republic of Lithuania, 2013).

⁶⁸ For example, see *Intermediate Report on Under-Resourced Languages*, European Commission, Brussels, 2017 // <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5aff37b1f&appId=PPGMS>.

⁶⁹ Aspect-based sentiment analysis is a type of text analysis that categorises opinions by aspect and identifies the sentiment (negative, positive, or neutral) related to each aspect.

extensive analysis of English legal language in many research papers and big datasets (labelled and unlabelled) available on open access. Also, advanced tools and solutions for English legal text analysis are open access and available on open access (most of them are still prototypes), e.g., the Blackstone project that contains the spaCy⁷⁰ model and a library for processing long-form, unstructured legal text.⁷¹ As indicated above, for Lithuanian, there is little linguistic legal language research, and there are no sufficient legal language resources (labelled and unlabelled) for machine learning needs. To fill those gaps, our research team has created a unique legal text corpus that contains a large sub-corpus of legal acts texts, taken from www.e-tar.lt, which is publicly available, along with a large sub-corpus of legal court decisions texts, taken from LITEKO, which is also publicly available, and a special sub-corpus that contains 1,500 anonymised real-life contracts (not available on open access). For research purposes, the contracts sub-corpus was synthetically augmented with fake data to make documents closer to real-life documents.

Predictive technologies are also lagging for Lithuanian and will continue to do so. For comparison, in the case of the English language, Washington University researchers successfully predicted the outcome of 75 percent of the Supreme Court's decisions during the 2002 term. This was in comparison with an expert panel reviewing the same material that predicted only 59.1 percent of case outcomes. Michigan State University researchers successfully predicted the outcomes of all Supreme Court cases from 1816 to 2015 with 70 percent accuracy⁷². The judicial decisions of the European Court of Human Rights (ECtHR) have been predicted with 79 percent accuracy using an AI method developed by researchers at the University of Sheffield and the University of Pennsylvania⁷³. These achievements are impressive, but those solutions still cannot be used for real-life needs, where expected accuracy must reach at least 95 percent. So, we agree with Hinton's opinion that conceptual breakthroughs in NLP are necessary, at least for under-resourced and small languages such as Lithuanian, to make use of AI and deep learning throughout the legal domain.

From our multi-aspect research of Lithuanian legal language and legal texts, we drew the following conclusions (this list is not exhaustive):

From *linguistic and lexical analysis*:

- The Lithuanian legal domain language differs a lot from the common language. In our legal corpus, unique words only account for about 2 percent of the words contained in the corpus. In our internet media corpus, unique words account

⁷⁰ See <https://spacy.io/>.

⁷¹ See <https://github.com/ICLRandD/Blackstone>.

⁷² See <https://www.exigent-group.com/blog/5-legal-technology-trends-to-watch-in-a-post-covid-world/>.

⁷³ See <https://www.ucl.ac.uk/news/2016/oct/ai-predicts-outcomes-human-rights-trials>.

for c. 8 percent; in our fiction-scientific corpus, c. 12 percent. We found the same percentages in the English legal domain corpora. Narrow vocabulary may indicate that legal domain documents are very suitable for state-of-the-art NLP algorithms because, to date, they have only used a very limited vocabulary (up to 30,000 words in BERT or c. 52,000 in RoBERTa). However, in legal documents, all semantics and legal reasoning depend on the context. As noted above, transformer-based models processing algorithms with attention mechanisms do well with context but only perform with legal documents because they are unable to process long sequences (the input is limited to 512 tokens). In our collection, a contract document contains, on average, 4,500 words. This means that documents must be split into subparts, thereby losing the context of the whole document.

- In Lithuanian, according to grammatical rules, the first letter of words is capitalised only in the case of proper nouns⁷⁴ and when words come at the beginning of sentences. In our contract subcorpus, we found c. 80,000 cases of words with capitalised first letters that were neither proper nouns nor at the beginning of sentences (e.g., "Client," "Contract," "Partner," etc.). This does not follow traditional grammar rules and makes commonly used information extraction tools useless (for example, morphological analysers, which take such words as proper nouns or named entity recognisers after being trained on online media texts). We found that the commonly used practice of using lowercase text in the preprocessing stage did not solve the problem; instead, it created new ones. We created a special syntactical parser for Lithuanian legal domain texts.

- Since word embedding and vectorised language models for the Lithuanian legal domain are not available and we had insufficient computational and language resources to create our own, for morphological tagging and disambiguation we used our own Hunspell-based morphological tagger, which we adjusted for legal domain vocabulary. With morphological tagging, we achieved 96 percent accuracy,⁷⁵ but the morphology and semantic disambiguation problem remains.

From *syntactical and form analysis*:

- Almost all legal documents have rigid structures and are well formatted. Generally, this is an aspect that fits well with machine learning algorithms' needs. However, this rule is not valid for legal documents. The specific format of legal documents is closely related to how logical units that contain semantic meaning and in general language are called sentences and are constructed. They do not follow the main language's syntactic rules. For this reason, common syntactic parsers are

⁷⁴ See <http://www.vlkk.lt/aktualiausios-temos/rasyba/terminu-rasymas-didziaja-raide>.

⁷⁵ The dictionary of our morphological analyser contains 180,000 lemmas (sing. nom.) of the most frequent Lithuanian words. According to the morphological rules of the Lithuanian language, from this number of lemmas we have obtained c. 20,000,000 morphological forms.

useless (including definitions of semantic roles in sentences). In legal documents, semantically meaningful structures are not always sentences in the traditional sense; notions such as preposition or clause are usually more suitable. Sometimes legal clauses contain a few sentences that are inseparable due to their complex meaning, but sometimes they do not contain all the syntactical elements they should. For example, syntactical constructions introduced by “Duties of the client are” are often followed by long lists of duty clauses, each presented on a separate line, introduced by a list number and an uppercase letter at the beginning, and with a comma or semicolon at the end of the line. They often lack a subject, and they are defined only at the beginning of the long list of duties.

- Since standardisation is lacking, there are so many forms of legal documents. Structure depends more on the specific court or other legal institution than on the type of document.

From *automated information extraction and data mining*:

- As noted above, the specific use of capitalised terms disrupts traditional patterns, making it difficult to recognise and extract named entities (persons, locations, etc.), which is an obstacle to automatically linking recognised entities. As there is a lack of data to train the implemented AI-based language processing model, there was an issue of how to introduce features to recognise and extract specific legal named entities automatically (actors of cases or contracts, etc.). Thus, for better results, we used two interconnected named entity recognisers—one neural method-based and the other, rule-based. We have achieved high levels of accuracy, but uncertainties always remain due to unseen cases (a weakness of rule-based methods). To solve these problems, we created a specific named entity recognition solution for the Lithuanian legal domain, which is not described in this paper.

- All legal text datasets that are available through open access are anonymised. To make automated entity linking and relation extraction possible, we artificially augmented our dataset with fake data that contained required additional human labour efforts and augmented financial expenses.

From *advanced AI and NLP*:

- To make use of all the advantages of AI and NLP, state-of-the-art technologies, especially pre-trained transformer-based models like the BERT family and GPT, and to have the option to create predictive systems, we must have really huge amounts of text data to train the models.⁷⁶ Google and Facebook have

⁷⁶ For example, BERT-large has 340 million parameters, 1,024 hidden layers, and is trained on a 3.4-billion word corpus. Although the larger model performs better, fine tuning and training such a model is difficult and requires a lot of computational horsepower. GPT-3’s full version has a capacity of 175 billion machine learning parameters and is trained on 499 billion words. GPT-3 also requires a lot of computational power to train and tune it.

produced real and large pre-trained models for Lithuanian. However, both of them were pre-trained on media texts, and thus they cannot be used in legal domain systems because the language patterns, vocabulary, semantics, etc. are different. In English, there have been attempts to create pre-trained models for the legal domain, but even in English, there are not enough texts to make really good models for the legal domain. For small languages, like Lithuanian, the number of documents is not sufficient. The only hope is that new AI solutions will require smaller sets of training data (advanced algorithms might converge well on smaller samples).

To sum up, automation of legal tasks for the Lithuanian legal domain is on the way, but the creation and adoption of LegalAI to its full extent for the Lithuanian legal domain are still in their early development stages. We believe that it is in the interests of both scientific progress and technological innovation to assume that this goal is achievable until proven otherwise. However, policymakers remain the decision makers—since Lithuanian is not a commercial language—in investing in the development of national language technologies for the legal domain or moving to English-language solutions.

CONCLUSIONS

In terms of technological advancement, the post-pandemic world is shifting from the age of digital electronics (Industry 3.0) to a society distinguished by the merging of physical and cyber spaces and the collaboration between machines and human beings (Industry 5.0),⁷⁷ where most current legal instruments are too slow and ineffective for regulating these transformations. AI technologies can affect legal processes in two broad ways: (a) by automation, which involves grafting new technology onto old working practices, i.e., improving, refining, streamlining, optimising, and turbo-charging our traditional work practices; and (b) by transformation, when technology displaces and revolutionises conventional working habits and brings radical change—doing new things, rather than old things in new ways. While automation is a reassuringly familiar practice, the dominant trend in the world of technology is towards transformation.

AI technologies can have an important role in improving access to justice, especially its procedural aspects, such as improving the *transparency of the proceedings, equality of arms, and cost of litigation*; however, this does not mean that the use of these technologies has facilitated better access to justice in practice now. On the contrary, the pandemic has revealed the extent of technological inequality, especially due to the lack of electronic devices or access to the internet.

⁷⁷ A.S. George and A.H. George, *supra* note 5.

When we are talking about improving access to justice, we are referring to much more than providing access to quicker, cheaper, and less complicated mechanisms for resolving disputes. We are also speaking of the introduction of techniques that deeply empower all members of society—to avoid disputes in the first place and, more importantly, to have greater insight into the benefits that the law can confer. Even very capable people today can feel disempowered when involved in legal processes. Tomorrow, we should want citizens to be able to own and manage many of their own legal issues. In our hyperregulated and legally reactive society, there is a need for profound legal system innovation and transformation, but this is not possible without real breakthroughs in AI, especially in its subfield, NLP.

Apart from NLP problems, it must be noted that existing AI tends to involve “narrow” intelligence, AI algorithms tend to work poorly, or not at all, in areas that are conceptual, abstract, value-laden, open-ended, or policy- or judgment-oriented, that require common sense or intuition, involve persuasion or arbitrary conversation, or involve engagement with the meaning of real-world humanistic concepts, such as societal norms, social constructs, or social institutions.

Empirical evidence shows that for small languages such as Lithuanian, there are major technical challenges that can hinder AI usage in the legal domain:

- Because there is insufficient data to train AI-based legal service models, the best-performing AI systems to date have been hybrid types—based on rules and neural methods. Neural methods are flexible, good at handling unseen data, and probabilistic, but they are also unpredictable and require a lot of training data. Rule-based methods are reliable, easy to interpret, require no training data, and are very predictable, but they are very narrowly defined and complex systems of rules that are difficult to understand/update/maintain, do not adapt well to unseen situations, and require a lot of hand coding work and the presence of experts in the field.

- Due to the problems described above, predictive technologies are still lagging for Lithuanian, and they will continue to do so.

- The situation is much better for legal process automation. In addition to the capacity to access and upload documents online and video/teleconference hearings—technologies that make access to justice easier⁷⁸—the vast majority of successful start-ups have good working solutions for automation tasks, such as document automation and legal case management (workflow management, billing and invoicing, reporting, and analytics). The results of our project can be added to

⁷⁸ According to the Access to Justice Scorecard 2020 survey by the Queensland Law Society, the vast majority of respondents agreed that technology has been identified as having an important role in access to justice, especially online access to information about laws and legal rights, the capacity to access and upload documents online, and video/teleconferencing of hearings. It must be noted that all these options require one detail—citizens must have access to the internet, and that is problematic in rural areas of Lithuania.

these achievements. In Lithuania, the e-service portal of the Lithuanian courts (also called "e-Court") has been launched, and it has moved a lot of justice administration online, including e-filing.⁷⁹ All those solutions and innovations are more efficient than their predecessors, which may make the lives of judges, attorneys, notaries, and other legal professionals easier and may reduce the costs of legal services.

BIBLIOGRAPHY

1. Amilevičius, Darius. "Machine Bias and Fundamental Rights": 334–365. In: John Gordon, ed. *Smart Technologies and Fundamental Rights*. Leiden: Brill, 2020.
2. Ashley, Kevin D. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge: Cambridge University Press, 2017.
3. Assy, Rabeeae. *Injustice in Person: The Right to Self-Representation*. Oxford: Oxford Scholarship Online, 2015.
4. Bard, Alexander, and Jan Soderqvist. *Syntheism: Creating God in the Internet Age*. Stockholm Text, 2014.
5. Bass, Julia, W. A. Bogart, and Frederick H. Zemans, eds. *Access to Justice for a New Century—The Way Forward*. Toronto: Law Society of Upper Canada, 2005.
6. Beltagy, I. "Longformer: The Long-Document Transformer." 2020 // <https://arxiv.org/pdf/2004.05150.pdf>.
7. Chalkidis, Ilias, et al. "Neural Legal Judgment Prediction in English." 2019 // <https://arxiv.org/pdf/1906.02059.pdf>.
8. Cui, Yadong. *Artificial Intelligence and Judicial Modernization*. Cham: Springer/Shanghai People's Publishing House. 2020.
9. European Parliamentary Research Service. *Common Minimum Standards of Civil Procedure, European Added Value Assessment*. Annex I. Strasbourg: European Parliament, 2016.
10. European Parliamentary Research Service. *Language Equality in the Digital Age*. Strasbourg: European Parliament, 2017 // <https://op.europa.eu/en/publication-detail/-/publication/fa0a50e7-cda4-11e7-a5d5-01aa75ed71a1/language-en/format-PDF>.
11. Francioni, Francesco, ed. *Access to Justice as a Human Right*. Oxford: Oxford University Press, 2007.
12. Freeman, Michal, and Fiona Smith, eds. *Law and Language*. Vol. 15. Oxford: Oxford University Press, 2011.

⁷⁹ See <https://e.teismas.lt/en/public/home/>.

13. George, A. S., and A. H. George. "Industrial Revolution 5.0: The Transformation of the Modern Manufacturing Process to Enable Man and Machine to Work Hand in Hand." *Seybold Report* 15, no. 9 (2020): 214–234.
14. Goodrich, Peter. *Legal Discourse*. New York: Palgrave Macmillan, 1987.
15. Hager, John W. "Let's Simplify Legal Language." *Rocky Mountain Law Review* 32 (1959): 74–86.
16. Hildebrandt, Mireille. *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology*. Cheltenham, UK: Edward Elgar, 2015.
17. Holder, Chris, Vikram Khurana, Faye Harrison, and Louisa Jacobs. "Robotics and Law: Key Legal and Regulatory Implications of the Robotics Age (Part I of II)." *Computer Law & Security Review* 32, no. 3 (June 2016): 383–402 // <https://doi.org/10.1016/j.clsr.2016.03.001>.
18. Kiršienė, Julija, Christopher Kelley, Deividas Kiršys, and Juras Žymančius. "Rethinking the Implications of Transformative Economic Innovations: Mapping Challenges of Private Law." *Baltic Journal of Law & Politics* 12, no. 2 (2019): 47–77.
19. Kiršienė, Julija, Edita Gruodytė, and Darius Amilevičius. "From Computerised Thing to Digital Being: Mission (Im) Possible?" *AI & Society* 36, no. 2 (2021): 547–60.
20. Li V., and S. la Roque-Doherty. "Toward Smarter Courts: Artificial Intelligence Has Made Great Inroads—But Not as Far as Increasing Access to Civil Justice." *ABA Journal* 107, no. 2 (April/May 2021) // <https://go.gale.com/ps/i.do?p=EAIM&u=anon~b1a1800&id=GALE|A655942226&v=2.1&it=r&sid=bookmark-EAIM&asid=e697e6f6>.
21. Manuel, Marcus, and Clare Manuel. "People-Centred Justice for All: A Route to Scaling Up Access to Justice Advice and Assistance in Low-Income Countries." 2021 // <https://odi.org/en/publications/people-centred-justice-for-all/>.
22. Marcus, Garry. "Deep Learning: A Critical Appraisal." 2018 // <https://arxiv.org/ftp/arxiv/papers/1801/1801.00631.pdf>.
23. McShane, M., and S. Nirenburg. *Linguistics for the Age of AI*. Cambridge, MA: MIT Press, 2021.
24. Mellinkoff, David. *The Language of the Law*. New York: Little, Brown and Co., 1963.
25. Meskys, Edvinas, Aidas Liaudanskas, Julija Kalpokiene, and Paulius Jurcys. "Regulating Deep Fakes: Legal and Ethical Considerations." *Journal of Intellectual Property Law & Practice* 15, no. 1 (2020): 24–31.

26. Mubarak, Usama. *Role of Legislative Drafting in Victimization of Citizens* (2020) // <http://dx.doi.org/10.2139/ssrn.3656529>.
27. OECD and Law & Justice Foundation of New South Wales. *Access to Justice and the COVID-19 Pandemic: Compendium of Country Practices* (Paris and Sydney: OECD and Law & Justice Foundation of New South Wales, 2020) // <https://www.oecd.org/governance/global-roundtables-access-to-justice/access-to-justice-compendium-of-country-practices.pdf>.
28. Philips, Alfred. *Lawyers' Language: How and Why Legal Language is Different*. New York, Routledge. 2003.
29. Queensland Law Society. "Access to Justice Scorecard." 2021 // <https://www.qls.com.au/Pages/Advocacy/Access-to-Justice-Scorecard>.
30. Schumpeter, J. A. *The Theory of Economic Development*. New Brunswick, NJ: Transaction Publishers, 1983.
31. Shikhar, Sanjana. "The Role of Artificial Intelligence in Law." 2021 // <https://blog.iplayers.in/role-of-artificial-intelligence-in-law/>.
32. Statal Commission for Lithuanian Language. *Guidelines for the Development of the Lithuanian Language in Information Technology 2014–2020*. Vilnius: Republic of Lithuania, 2013.
33. Susskind, Richard. *The Future of Law: Facing the Challenges of Information Technologies*. Oxford: Clarendon Press. 1998.
34. Turing, Alan. "Computing Machinery and Intelligence." *Mind* 59, no. 236 (1950): 433–460.
35. United Nations. "World Social Report 2020: Inequality in a Rapidly Changing World" // <https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/01/World-Social-Report-2020-FullReport.pdf>.
36. Vladeck, David C. "Machines Without Principals: Liability Rules and Artificial Intelligence." *Washington Law Review* 89, no. 1 (March 2014): 117–150.
37. Waelbers, Katinka, and Tsjalling Swierstra. "The Family of the Future: How Technologies Can Lead to Moral Change": 219–236. In: Jeroen van den Hoven, Neelke Doorn, Tsjalling Swierstra, Bert-Jaap Koops, and Henny Romijn, eds. *Responsible Innovation 1*. Dordrecht: Springer Netherlands, 2014 // https://doi.org/10.1007/978-94-017-8956-1_12.
38. World Economic Forum. "Values, Ethics and Innovation: Rethinking Technological Development in the Fourth Industrial Revolution" // <https://www.weforum.org/whitepapers/values-ethics-and-innovation-rethinking-technological-development-in-the-fourth-industrial-revolution/>.
39. World Justice Project. "Global Insights on Access to Justice 2019" //

<https://worldjusticeproject.org/our-work/publications/special-reports/global-insights-access-justice-2019>.

40. Xiao, Chaojun, *et al.* "Lawformer: A Pre-trained Language Model for Chinese Legal Long Documents." 2021 // <https://arxiv.org/pdf/2105.03887.pdf>.
41. Zhong, Haoxi. "How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence." 2020 // <https://arxiv.org/abs/2004.12158>.

LEGAL REFERENCES

1. *Common Minimum Standards of Civil Procedure*. European Parliament. European Parliamentary Research Service. European Added Value Assessment, Annex I. Strasbourg, 2016.
2. *Development Program of the Justice System in Lithuania (2021–2030)*. Government of the Republic of Lithuania // <https://tm.lrv.lt/uploads/tm/documents/files/dokumentai/Administracin%C4%97%20informacija/PI%C4%97tros%20programos/2021-2030%20metu%20teisingumo%20sistemas%20pletros%20programa.pdf>.
3. *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. European Commission for the Efficiency of Justice (CEPEJ). Brussels, 2018 // <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.
4. *Intermediate Report on Under-Resourced Languages*. European Commission. Brussels, 2017 // <https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5aff37b1f&appId=PPGMS>.
5. *Presidency Conclusions—The Charter of Fundamental Rights in the Context of Artificial Intelligence and Digital Change*. Council of the European Union. 11481/20. Brussels, 2020.
6. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. European Commission. COM(2021) 206 final, 2021/0106. Brussels, 2021.
7. *Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*. European Commission. Brussels, 2021.

8. *Resolution of 20 October 2020 on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies*. European Parliament. 2020/2012(INL). Strasbourg, 2020.
9. *Universal Declaration of Human Rights (UDHR)*. United Nations General Assembly, New York, 1948.